IJCSIS Vol. 9 No. 9, September 2011
ISSN 1947-5500

International Journal of Computer Science & Information Security

Editorial
Message from Managing Editor

*The Journal of Computer Science and Information Security (**IJCSIS**) offers a track of quality R&D updates from key experts and provides an opportunity in bringing in the new techniques and horizons that will contribute to advancements in Computer Science in the next few years. **IJCSIS** scholarly journal promotes and publishes original high quality research dealing with theoretical and scientific aspects in all disciplines of Computing and Information Security. Papers that can provide both theoretical analysis, along with carefully designed computational experiments, are particularly welcome. **IJCSIS** is published with online version and print versions (on-demand).*

***IJCSIS** editorial board consists of several internationally recognized experts and guest editors. Wide circulation is assured because libraries and individuals, worldwide, subscribe and reference to **IJCSIS**. The Journal has grown rapidly to its currently level of over thousands articles published and indexed; with distribution to librarians, universities, research centers, researchers in computing, and computer scientists. After a very careful reviewing process, the editorial committee accepts outstanding papers, among many highly qualified submissions. All submitted papers are peer reviewed and accepted papers are published in the IJCSIS proceeding **(ISSN 1947-5500**). Both academia and industries are invited to present their papers dealing with state-of-art research and future developments. IJCSIS promotes fundamental and applied research continuing advanced academic education and transfers knowledge between involved both sides of and the application of Information Technology and Computer Science.*

The journal covers the frontier issues in the engineering and the computer science and their applications in business, industry and other subjects. (See monthly Call for Papers)

Since 2009, ***IJCSIS** is published using an open access publication model, meaning that all interested readers will be able to freely access the journal online without the need for a subscription. On behalf of the editorial committee, I would like to express my sincere thanks to all authors and reviewers for their great contribution.*

Available at http://sites.google.com/site/ijcsis/

IJCSIS Vol. 9, No. 9, September 2011 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

(1) Superintending Engineer ONGC Research scholar Punjabi university Patiala, India
(2) Director, University College of Engineering Punjabi University Patiala, India

23. Paper 31081140: PAPR Reduction of OFDM Signal Using Error Correcting Code and Selective Mapping (pp. 136-139)

Anshu, ECE Deptartment, Mahrishi Markendeshwar University, Mullana, India
Er. Himanshu Sharma, Lecturer, ECE Department, Maharishi Markendeshwar University, Mullana, India

24. Paper 31081143: Lossless Image Compression for Transmitting over Low Bandwidth Line (pp. 140-145)

G. Murugan, Research Scholar, Singhania University & Sri Venkateswara College of Engg, Thiruvallur
Dr. E. Kannan, Supervisor, Singhania University and Dean Academic Veltech University
S. Arun, ECE Dept. Asst. Professor Veltech High Engg college, Chennai

Using Image Steganography to Establish Covert Communication Channels

Keith L Haynes
Center for Security Studies
University of Maryland University College
Adelphi, Maryland, USA
keith.haynes@earthlink.net

Steganography is the art or science of sending and receiving hidden information. This paper investigates the use of image steganography to breach an organization's physical and cyber defenses to steal valuable information. Furthermore, it proposes a steganographic technique that exploits the characteristics of the computer vision process that are favorable for encryption. The result is an image steganographic system that is undetectable and secure.

Keywords- Steganography, computer vision, machine learning, image hiding

I. INTRODUCTION

Steganography is a form of secret communication that has been in existence for thousands of years. One of the earliest examples occurred around 440 BC and was noted in an ancient work entitled "Histories of Herodotus." Herodotus recounts how Histiæus shaved the head of his most trusted slave and tattooed it with a message to instigate a revolt against the Persians. The message was covered when the slave's hair regrew [5]. With the advent of digital technology, there has been considerable effort placed in finding effective means of hiding data in digital media; photo images in particular. However, if the hidden message is discovered, its information is compromised. Encryption, on the other hand, does not seek to hide the information; rather it encodes the information in such a fashion that it appears meaningless to unauthorized observers. If an encrypted data stream is intercepted and cannot be decrypted, it is still evidence that secret communication is occurring and may compromise the sender or the receiver. An ideal form of secret communication would combine the hidden aspect of steganography with a strong cryptographic algorithm.

The Internet has evolved into a media rich environment with countless numbers of photographic images being posted to websites or transmitted via email every day. Thus, digital images provide an excellent cover for covert communications because their presence on the Internet does not draw significant attention, other than their visual content. This should be of concern to security personnel because it opens the possibility of undetectable lines of communication being established in and out of an organization with global reach. "Computer hacking is not a new crime, nor is insider trading, but the Securities and Exchange Commission (SEC) has recently focused its attention on computer hackers trading on wrongfully obtained inside information." [9] Image steganography can be utilized to facilitate this type of crime. For example, an employee of a large corporation could update his/her Facebook page with vacation photos that contain hidden insider trading or other sensitive information. The message does not have to be long. A message as simple as "sell stock" or "buy stock" can be quite effective. In general, "there are five steps to follow to carry out a successful cyber-attack: find the target; penetrate it; co-opt it; conceal what you have done long enough for it to have an effect; and do something that can't be reversed." [10] Steganography aids in the concealment of these illegal activities by providing covert communication channels.

This paper proposes a novel method for image Steganography that represents a major departure from traditional approaches to this problem. This method utilizes Computer Vision and Machine Learning techniques to produce messages that are undetectable and if intercepted; cannot be decrypted without key compromise. Rather than modify the images, the visual content of the images is interpreted from a series of images.

A. Motivation

Numerous methods of Steganography have been proposed that utilize images as covers for secret messages. These methods fall into three main categories [1]:

- Least Significant Bit (LSB) – encodes a secret message into an existing image by modifying the least significant bits of pixel [11].
- Injection – utilizes the portion of the image file that is not required for rendering of the image to write the hidden message.
- Substitution – is similar to LSB, but attempts to minimize distortion caused by changing pixel values. A simple LSB substitution, which hides data into LSBs directly, is easily implemented but will yield a low quality stego-image. In order to achieve a good quality stego-image, a substitution matrix can be used to transform the secret data values prior to embedding into the cover image. However, there can be difficulty in finding a suitable matrix.[12]

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

LSB, injection, and substitution methods all use an original or cover images to created stego-images that contain the hidden messages. The steganographic process usually begins with the identification of redundant bits in the cover image and replacing those bits with bits from the secret message. The modification of the image leaves minor distortions or detectable traces that can be identified by statistical steganalysis. In an effort to avoid detection, many varied techniques have been proposed. Recently, Al-Ataby and Al-Naima proposed an eight step method that utilizes Discrete Wavelet Transform (DWT) [6]. The first step in the process tests the cover image for suitability. If the cover image is acceptable, it is processed prior to encoding. An encryption cipher is required to protect the secret message. In the final steps, the encrypted message and processed cover image are combined; forming the stego-image. The process suffers from two problems. First, the criteria for the cover images limit the amount of images that can be utilized. Secondly, though the process is less susceptible to statistical steganalysis, however, since the cover image is modified, comparison with the original image may reveal the presence of manipulated data. There are cybersecurity countermeasures that can be employed to protect against the threat that procedures such as this can present. Gutub et al. proposed a pixel indicator technique, which is a form of steganographic substitution [14]. The method utilizes the two least significant bits of the different color channels in an RGB scheme. The bits are used to indicate the presence of secret data in the other two channels. The actual indicator color channel used is randomly set based to the characteristics of the images. Because of the fact that the image is modified, it is vulnerable to the same attacks as other LSB or other substitution methods.

Techniques have been proposed to remove steganographic payloads for images. Moskowitz et al. proposed one such method that utilized what they called an image scrubber [2]. In order for the image scrubber to be effective in preventing image steganographic communications, it must be applied to all images traversing the organization boundary. Additionally, it must not distort the visual information contained in the image file because most of the digital images transmitted are valid files and not stego-images. Methods like image scrubbing and other forms of steganographic cryptanalysis can be effective on the aforementioned techniques; however, they would fail if a technique employed was based on the informational content of unmodified images. Since Computer Vision is not normally associated with steganography and encryption, the next section will provide a brief introduction for readers who are not familiar with its fundamental concepts.

II. COMPUTER VISION BACKGROUND

In essence, computer vision is the science and technology that allow machines to see. More specifically, the goal of a vision system is to allow machines to analyze an image and make a decision as to the content of that image. That machine-made decision should match that of a human performing the same task. An additional goal of a vision system is to identify information contained in an image that is not easily detectable by humans. As a science, computer vision is still in its infancy; however, there are many applications in existence, such as, automatic assembly lines using computer vision to locate parts. The two primary computer vision tasks are detection - determining whether an object is present in an image and recognition - distinguishing between objects. Most computer vision systems fall into two main categories: Model-Based or Appearance-Based. Model-Based computer vision relies on the knowledge of the system's designer to create 3D models of the objects of interest to be used by the system for comparison with image scene. Appearance-Based systems, on the other hand, use example images and machine learning techniques to identify significant areas or aspects of images that are important for discrimination of objects contained within the image.

A. *Machine Learning*

A key aspect of machine learning is that it is different from human knowledge or learning. This difference is exemplified by the task of face detection. A child is taught to recognize a face by identifying the key features such as eyes, nose, and mouth. However, these features do not exist in the context of machine learning. A computer has to make a decision of the presence of a face based on the numbers contained in a 2D matrix such as the one in Figure 1. The matrix contains the grayscale pixel values for a 24 X 24 image of a face. The matrix highlights two aspects that make computer vision a very difficult problem. First, humans do not possess the ability to describe the wide variety of faces in terms of a 2D numerical matrix. Secondly, analysis of the photographic images involves handling extremely high dimensional data; in this case, the face is described by a vector of 576 values. This problem is known as the "Curse of Dimensionality" [3]. In short, as the dimensions increase, the volume of the space increases exponentially. As a result, the data points occupy a volume that is mainly empty. Under these conditions, tasks such as estimating a probability distribution function become very difficult or even intractable.

15	15	32	44	57	84	138	219	244	248	248	248	248	246	244	242	223	222	233	244	245	223	160	74
9	14	36	50	57	81	119	128	208	244	250	248	251	221	153	145	158	191	209	228	217	177	133	62
36	27	54	87	106	121	149	169	133	126	160	222	226	171	150	182	177	175	176	179	172	158	122	35
27	56	100	124	144	155	144	147	86	42	64	165	190	152	188	212	173	162	187	198	196	174	110	40
11	69	97	97	105	112	91	80	46	15	41	157	186	146	182	160	113	100	152	188	202	188	119	40
22	52	41	31	29	28	36	43	5	2	52	173	187	122	135	79	47	19	52	90	131	168	142	35
29	33	20	19	23	27	46	69	52	11	33	146	174	115	99	48	35	31	31	52	97	148	150	74
17	41	48	72	104	122	153	237	235	56	33	162	242	175	73	91	113	152	181	197	201	192	167	134
17	41	74	80	76	75	106	224	235	51	36	165	251	183	71	120	103	136	194	208	199	195	171	130
29	54	94	107	101	94	122	212	119	30	46	168	251	225	167	148	141	125	175	190	180	176	154	116
44	72	93	100	104	113	111	80	33	11	43	163	242	228	182	108	163	157	156	143	150	166	141	107
50	99	142	126	108	110	79	10	52	37	54	166	243	229	194	140	163	157	155	147	140	132	111	94
43	103	161	165	158	160	116	20	97	84	81	173	244	234	215	200	178	160	165	166	147	120	102	94
33	84	142	191	224	234	185	53	125	110	76	160	240	223	194	211	202	184	171	164	154	137	119	109
35	76	165	222	243	230	159	73	127	101	50	139	230	201	155	195	189	183	171	171	160	139	128	122
42	89	186	230	231	177	62	25	27	61	100	159	196	191	178	167	135	153	165	183	173	146	141	126
50	138	191	173	138	97	35	10	10	42	83	142	166	131	83	56	68	71	136	187	192	176	154	108
38	133	116	83	78	64	29	14	15	61	119	182	189	135	82	51	50	54	66	148	198	184	167	112
28	96	89	81	93	81	37	22	69	109	157	190	203	196	171	148	74	67	49	107	167	179	167	93
26	77	127	157	160	114	34	13	77	150	200	209	215	229	224	197	52	40	68	94	129	165	151	70
22	60	159	210	191	126	44	19	40	101	145	152	161	173	164	151	76	94	145	156	155	158	122	41
14	33	134	187	170	122	71	47	33	53	91	106	125	144	131	140	171	207	227	232	207	154	86	12
6	18	74	122	143	128	85	71	77	113	164	185	204	226	225	227	235	234	239	235	196	125	49	1
0	12	39	67	111	131	95	95	96	121	127	168	212	224	225	232	245	241	245	243	175	72	19	0

Figure 1. Greyscale Bitmap

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The Machine Learning approach to solving this problem is to collect a set of images that relate to the particular task to be performed. For face detection, two sets or classes of images are needed: one containing faces and one containing non-faces. These two sets form the training set. Note that the dimensions of all of the images in the training set should be approximately the same. Next, the designer must identify the type of features that will be used for image analysis. A feature is a calculation performed on a section of the image that yields a numerical value. The simplest feature is a pixel value; however, because of the number of pixels in an image and the high degree of variability between subjects, they are not often used directly as features. Instead, a feature is usually a summary computation such as an average, sum, or difference performed over a group of pixels. By summarizing key areas, the dimensionality of the problem is reduced from the number of pixels in the image to a much smaller set of features. An example of a feature is a Haar feature. A Haar feature is a number that is the result of the difference of two or more adjacent rectangular areas. The use of this type of feature in computer vision application was described by Papageorgiou et al [8]. Figure 2 shows five different Haar features. The sum of the pixels in the grey area is subtracted from the sum of the pixels in the white area. Note that Haar features are just one of many types of features that can be used. Any valid calculation on pixels that yields a number is suitable; therefore, the magnitude of the set of possible features is infinite. Finally, type 0 in Figure 2 is not a true Haar feature. It is simply the average over the range of pixels.

With the feature type has been identified, the actual machine learning process can begin. The goal of the process is to identify the set of features that "best" distinguishes between images in the different classes. The actual metric that defines what is meant by "best" must be established. It could be as simple as recognition accuracy. The metric used in this paper is called the F statistic [4] and defines how well the classes are separated from each other in the feature space; the details of this metric go beyond the scope of this paper. Since the "best" features are not known, an exhaustive search of all possible features of the chosen type is performed in a systematic manner. Haar features are rectangular; therefore, all of the possible rectangles in the image are evaluated. The image in Figure 1 is a 24 X 24 bitmap and 45396 rectangles can be found within the image. Since there are five types of Haar features used in this example, there are 222980 possible Haar features in an image of this size. Each rectangular feature is applied one at a time to all of the images in the training set. The feature that best separates the classes is selected. Normally, one feature is insufficient to accurately distinguish between the classes; therefore, another search is conducted to find a feature that works best in conjunction with the first feature. This process is continued until an acceptable level of accuracy is achieved [13].

B. Feature Space

Once the feature set has been determined, a mapping of the solution between features and classes can be created. This mapping is generated by traversing the space defined by the features and labeling the class found at the various locations.

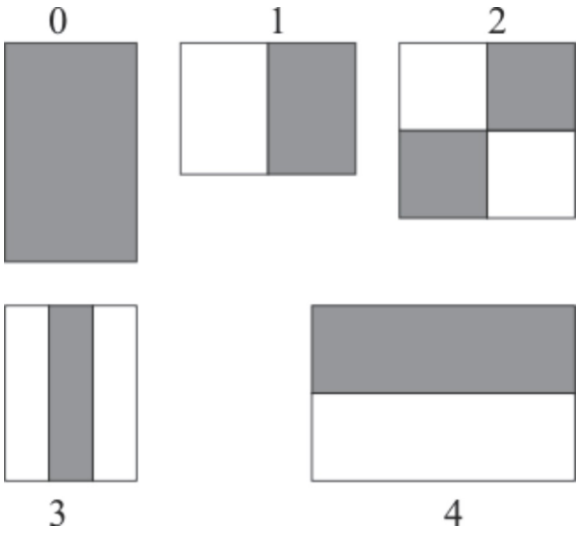


Figure 2. Haar Features

A feature set for a computer vision problem can contain a large number of features which define a high dimensional hyperspace with the same number of dimensions. Figure 3 depicts a 2D example of a feature space consisting of ten classes and two features. It also contains the solution of a nearest neighbor classifier derived from the initial feature space. The horizontal axes of each space represents the valid values for feature 1, similarly the vertical axes represent the valid values for feature 2. The different colors in the figure represent ten different classes for the problem. In this case, the two features effectively cluster images within the same class and provide separation between the different classes. As a result, the nearest neighbor classifier derived from this feature space is well-behaved and should yield a high accuracy level.

On the other hand, Figure 4 depicts a case where the two features do not effectively separate the classes. The result is a chaotic space where the classes are intermingled resulting in a low level of recognition or detection accuracy by the classifier. Note that if the training set or features used are changed, the feature space will be changed.

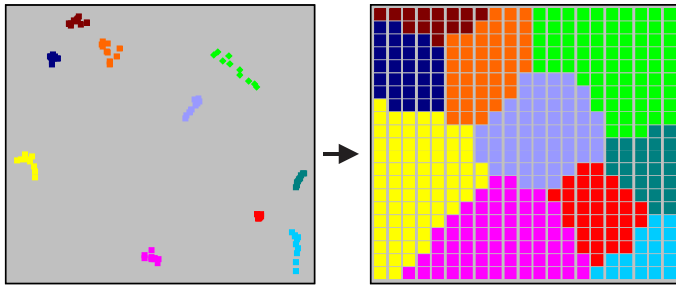


Figure 2. Feature Space and Solution Space

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

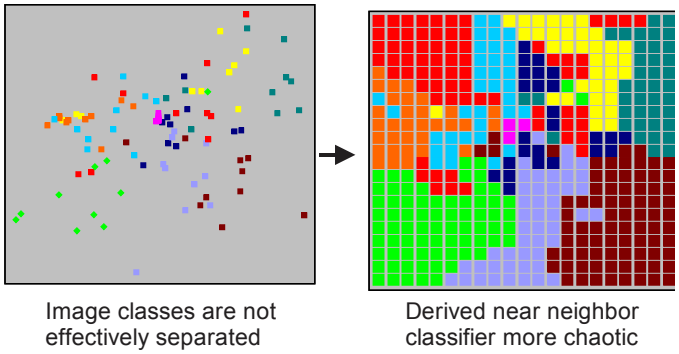


Figure 4. Poorly Separated Classes

C. *Classification Process*

With the classifier complete, the detection or recognition process is straightforward:

- Perform Feature Extraction on the target image. In other words, perform the calculations specified by the feature set in the image. The result is a vector of numerical values that represent the image.
- The vector is used as input into the classifier created from the feature space. The classifier determines the class contained in the image based on its solution space.

III. PROPOSED METHOD

The proposed method differs from other image Steganography methods in that the cover image does not contain a secret message; rather the classification of the image yields the hidden message. The algorithm is as follows:

1. Identify the characters that will be used to form the alphabet for communication.
2. Create a training set with the numbers of classes equal to the number of characters in the alphabet.
3. Use the training set to create a classifier using a Machine Learning process.
4. Collect a large number of images to be used to create messages and using the classifier, assign the collected images to classes.
5. Create a message by selecting images from the appropriate classes. The message can be transmitted by posting the images to a web page or sent via email.
6. Decode the message using the same classifier and class to character mapping.

A. *Alphabet Selection*

The selection of a suitable alphabet is a key step in this process. A generic alphabet that consists of all of the letters in the English alphabet, digits from 0 to 9, special characters such as a space can be utilized. The problem with an alphabet of this type is that steganographic messages formed would require numerous images to transmit simple messages. A better approach is to form an alphabet using words or phrases that relate to the type of data being communicated. Referring back to the insider trading scenario mentioned earlier, instead of spelling out buy, sell, or stock, the alphabet should contain a single character for each of the words. Using this alphabet, the message "sell stock" would require only two characters instead of 10. Once the alphabet is set, the number of classes needed in the training set is also fixed. It should be noted that this is not a high bandwidth procedure; however, there are many covert situations that require only a few words to be effective and have devastating effects.

B. *Training Set Creation*

The training set is the collection of images that will be used to determine the feature space. In normal vision systems, a small number of images (four – six) are assigned to each of the classes in the problem. The images assigned to a single class are related. The goal of the process is to yield a "well-behaved" feature space such as the one in Figure 1 that can accurately distinguish members of the different classes. However, in this system, unrelated images are arbitrarily assigned to the classes in the training set. The feature space generated from this type of training set will be chaotic. Moreover, the feature space will be unique for each training set formed. An important point that must be highlighted is that images used can be acquired from any source or several sources; therefore, the system can take advantage of the plethora of images available on the Internet and other sources. The only restriction is a minimum height and width dictated by the features used in the next step.

C. *Classifier Training*

Choosing a type of feature and classifier is the critical step in this process. It is important to note that since the goal is not to perform an actual computer vision task, accuracy is not desired. Since accuracy is not desired, any type of local feature or machine learning method can be used; however, there are desirable attributes. It does not matter what class an image is assigned to as long as the classification is consistent. Additionally, the generated feature space should be discrete consisting of bins or subregions. This attribute will allow the overall procedure to be resistant minor changes in the image file that may occur if the image is modified by cybersecurity measures. This attribute is depicted by the squares that make up the feature spaces in Figure 1.

As stated earlier, any suitable feature and classifier pairing can be used, however, the pairing utilized in this paper consist of Haar features and a Rapid Classification Tree (RCT). The details on the training process and use of this pairing are discussed in "Object Recognition Using Rapid Classification Trees [4]. Normally, the training process terminates when the selected feature set

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

can achieve a predetermined level of accuracy on the training set. Being that the training set contains a collection of arbitrary images, a high level of accuracy will not be achieved; and therefore, the desired number of features selected by the process should be specified prior to the training process. A feature set containing five or more features should provide sufficient security due to the complexity of the feature space it defines.

The steganographic method proposed in this paper is a form of symmetric key encryption because the same feature extractor and classifier is used for both encryption and decryption. The feature set and feature space form the key and are where the cryptographic strength of the process lies and is the only part of the process that must be kept secret. Furthermore, since the images are not modified, there is no evidence in the steganographic message that can be used to deduce the key. Without compromise of the key, encrypted messages will not be cracked. This point will be discussed in more detail in the discussion section of this paper. Once the classifier is completed, it can be shared with the members of the communication circle.

D. Image Collection

Once the classifier is trained, images must be collected and sorted into classes. As with the training set, the images that will be used to transmit messages can be acquired from any source. This fact makes the method a significant threat to cybersecurity. First, nearly all available images are suitable for the process; therefore, once a communication channel is established there is an endless supply of images for messages. Secondly, the visual content of the images can be used to hide the covert activity, by using themes. A website about baseball, sewing, celebrities could be used as a cover to transmit secret information globally. Finally, the abundance of images allows for images to be used only once. If no images are reused, the process is equivalent to a one-time pad, which is provably unbreakable [7].

Before they can be used, the images must be assigned to the various classes. With the trained classifier, this is a relatively simple task. Because of the chaotic nature of the feature space, all classes will be populated as long as a sufficient number of images are collected. It is important to note that the collection of images is not a one-time event; the supply of images can be replenished repeatedly.

E. Creating, Transmitting, and Receiving Secret Messages

Messages are assembled by selecting images from classes that correspond to the characters required to complete the message. The order of the images is maintained by naming the selected images in alphabetical or numerical order. Once the images are selected and ordered, the message can be assembled and transmitted. A serious threat to cybersecurity is posed by the fact that messages can be transmitted by various means to include copying to a thumb drive, attaching to an email, or posting to a webpage. A webpage poses a more significant threat because the number of recipients is not limited and are anonymous; unlike email where the recipients are identified.

Receiving a message is relatively simple. The image files must be received or downloaded to a system with the trained classifier. Once downloaded, the images are classified, thus revealing the associated characters. Note that the trained classifier does not require significant computing power. In fact, a handheld PDA with the classifiers software successfully decoded steganographic messages posted to a web page in less than 10 seconds. This fact poses another significant threat to cybersecurity; virtually any internet-enabled device can exploit this procedure. Therefore, traditional network security devices can be bypassed using the cellular network.

IV. Experimental Results

In order to demonstrate the procedure and its effectiveness, 984 images were arbitrarily collected from the Internet. The minimum size for the images was 128 by 128 pixels. This does not mean that the images were 128 by 128; but that image had a width or height below 128. The minimum size determines the number of rectangular areas in the images that can be used for features. In this case, there are 66,064,384 different rectangles that can be used for features. Fifty classes were used in this implementation. A training set was constructed by randomly distributing 4 images to each of the classes. The machine-learning process was run [4] and yielded the 10 Haar features depicted in Figure 5. The figure shows only the areas and type of feature used for determining the class of each of the images. The rectangles show the location and the color represents the type of filter used.

An i7 desktop computer with 8 GB of RAM was used for the experiment. Using this system the feature search took only 50 seconds and the final class recognition was only 11.5%. A nearest neighbor classifier was created using the results of the feature search. The remaining images were sorted into the proper classes using the classifier. Note that original feature search that took 50 seconds is the time consuming part of the process; the actual classification of an image is quick.

V. Discussion

It was asserted earlier in the paper that the system was undetectable and unbreakable without key compromise. In reference to the detectability, this process uses unmodified images that can come from any source. There is insufficient evidence to point to any covert communication because images traversing the Internet are commonplace. There is nothing to distinguish between a normal email and one containing a message.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

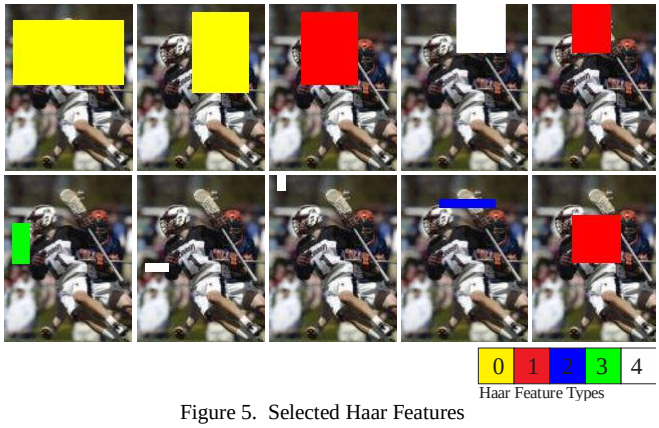


Figure 5. Selected Haar Features

Similarly, a message created using this method is unbreakable because the message provides insufficient evidence of the enormous complexity of the encryption. Suppose four steganographic messages and their corresponding plaintexts were capture. Additionally, all images come from the same class and representing the word "buy." Figure 6 contains the four images that were captured. The images belong to the same class because the classifier classified the as such. The first problem facing someone trying the defeat the system is identifying the features that are being used for classification. The message provides no evidence to solve this part of the problem. The entire image is not used for classification purposes, only the designated regions shown in Figure 5. Haar features are not the only type of features that could be used; any valid calculation on a set of pixel can be used as a feature. Assuming that the type of feature used is known, the problem is still too large to handle. Remember that a 128 by 128 image contains 66,064,384 different rectangles subregions and with the use of five different types of Haar features there are 330,321,920 possible features in a single image. However, the problem is still more complicated, because classification is based on a set of features; not a single feature.

The set can contain one, two, ten, or more different features. Again evidence is lacking to indicate what feature set is being used. When number of possible feature sets is consider, the magnitude of search space increases to 1.5466 X 1085; a space too large for a brute force attack. The classification computations performed on the four captured images in Figure 6 are not based on the images directly, but rather on the four row vectors contained in Table 1. Without the correct set of features, the vectors representing the images cannot be derived.

Figure 5. Selected Haar Features

TABLE I. Feature Values.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
106.5	110.6	-19.9	47.9	-4.8	1.4	1.8	48.6	3.6	-31.6
96.1	95.5	-5.8	0.1	1.4	-1.5	21.3	17.7	-3.2	-23.7
84.2	77.2	4.4	-5.8	-10.6	-16.2	10.8	-1.5	39.1	19.7
132.8	134.8	-4.0	-1.8	-3.7	16.5	14.2	-3.3	-29.9	12.5

If by chance one could determine the ten-feature set, it would only provide the inputs to classifier. The ten dimensional space define by the classifier is still unknown and massive. The row vectors in Table one equate to only four points in the space. Because of the chaotic nature of the feature space and that images are not reused, it is unlikely that new messages will map to known points in the space.

Finally, an effective feature search cannot be performed not only because of the massive size of the space that needs to be searched, but because there is no clear stopping to signal when the correct feature set is found. Table 2 contains the relative position of the values in Table 1 in the overall feature space. Zero percent would represent a feature value that is at the minimum of the range of values for that feature, while a value of 50% would be exactly halfway through the range. As the relative values are examined, it becomes clear that the values are not clustered. Therefore, as this search of possible feature sets there is no clear indication when the correct set has been found. Again, the images do not provide sufficient evidence to assist in analysis of the message. To further emphasize this point the transmitted images were all in color; however, all of the analysis was done in grayscale.

TABLE II. Feature Relative Position

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
31%	30%	38%	80%	46%	25%	61%	90%	64%	32%
25%	22%	45%	39%	50%	23%	81%	67%	59%	36%
20%	12%	51%	34%	43%	15%	70%	53%	91%	61%
44%	42%	46%	38%	47%	32%	74%	52%	40%	57%

VI. CONCLUSION

The method discussed in this paper represents a significant departure from traditional methods of image steganography; however, more significantly it poses a serious significant threat to any organization's cybersecurity. Because it utilizes ordinary unmodified images, there are no inherent indicators of covert communication taking place. The complexity of the encryption is such that without the key, transmitted messages will be secure. Finally, the small computational overhead, allows the method to be used

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

by virtually any Internet-enabled device to include cell phones; thus, creating many possible channels for secret communication.

ACKNOWLEDGMENT

I would like to thank Dr. Amjad Ali for his guidance and patience. Additionally, I would like to thank Dr. Jason O'Kane and my wife Levern for all of their help. Finally, a special thanks to Dr. Frank J. Mabry for introducing me to Steganography.

REFERENCES

[1] Narayana, S. & Prasad, G., (2010) "Two new Approaches for secured image steganography using cryptographic techniques and type conversion", Signal & Image Processing: An International Journal, Vol. 1, Issue 2, pp60-73.

[2] Moskowitz, I., Ahmed, F., & Lafferty, P., (2010), Information theoretic effects of JPEG compression on image steganography, International Journal of Computers & Applications; 2010, Vol. 32 Issue 3, pp318-327.

[3] R. E. Bellman, Adaptive Control Processes, Princeton University Press, Princeton, NJ, 1961.

[4] Haynes K, Liu X, Mio W, (2006), Object recognition using rapid classification trees, 2006 International Conference on Image Processing

[5] Katzenbeisser, S., Petitcolas, F (1999), Information Hiding Techniques for Steganography and Digital Watermarking, Artech House Books

[6] Al-Ataby, A., Al-Naima, F., (2010), A modified high capacity image steganography technique based on wavelet transform, The International Arab Journal of Information Technology, Vol. 7, Issue 4

[7] Shannon, C, (1949), Bell Labs Technical Journal in 1949, Bell Labs

[8] Papageorgiou, C., Oren, M., and Poggio, T. A general framework for object detection. In International Conference on Computer Vision, 1998

[9] Denny, R., 2010, Beyond mere theft: Why computer hackers trading on wrongfully acquired information should be held accountable under the securities exchange act, Utah Law Review; Vol. 2010, Issue 3, p963-982

[10] Creeger, M., 2010, The theft of business innovation: An ACM-BCS roundtable on threats to global competitiveness, Communications of the ACM, Vol. 53, Issue 12, p48-55

[11] Van Schyndel, R., Trikel, A., and Osborne, C., "A digital watermark" Proceedings of IEEE International Conference on Image Processing, Vol 2, pp. 86-90

[12] Yang, C., and Wang, S., 2010, Transforming LSB substitution for image-based steganography in matching algorithms, Journal of Information Science & Engineering, Vol. 26, Issue 4, p1199-1212

[13] Duda, R., Hart, P., and Stork, D., 2001, Pattern Classification, John Wiley & Sons, Inc, New York

[14] Gutub, A., Ankeer, M., Abu-Ghalioun, M., Shaheen, A., and Alvi, A., 2008, Pixel indicator high capacity technique for RGB image based steganography, WoSPA 2008 – 5th IEEE International Workshop on Signal Processing and its Applications

AUTHORS PROFILE

Keith Haynes received his BS in Chemistry from Hofstra University in Hempstead, NY. He then received a MS in Systems Management from Golden Gate University in 1990. He attended the Naval Postgraduate School and in 1993 received two MS degrees in Computer Science and Engineering Science. In 2006, he completed his PhD in Computer Science at Florida State University.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Virtual Education and its Importance as a New Method in Educational System

Mohammad Behrouzian Nejad
Young Researchers Club, Dezfoul Branch
Islamic Azad University
Dezfoul, Iran
Mohamad.behrouzian@gmail.com

Ebrahim Behrouzian Nejad
Department of Computer Engineering
Shoushtar Branch, Islamic Azad University
Shoushtar, Iran
E.behrozian@iau-shoushtar.ac.ir

***Abstract*— The increasing development of technology, especially information technology in education has led to many changes, including the cases that can be pointed to the emergence of Virtual Education. Virtual Education have been affected teaching and learning systems and itself as one of the main methods of learning has emerged. Courses offered in the multimedia environment removing the limitations of time and place for inclusive education to provide rapid feedback and such cases the advantages of this method is one of education. In the near future other structure and process of traditional training needs of human society not responsive in the information age, but knowledge is central to the goal. So Virtual Education as a new method and efficient can be very useful. In this paper we will examine the concepts, advantages, features and differences between traditional learning and teaching quality and efficiency to help executives implement effective this training method which can commensurate with the circumstances which they are located and make correct decisions in the application, implementation and development of Virtual Education.**

Keywords- Virtual Education, E-Learning, Educational Technology; Information Technology.

I. INTRODUCTION

Today, education is known as basic human rights, social progress and change agent. World of today is the world of science, knowledge and progress in any society and is based on the information. With the development of IT and telecommunications equipment to the depth of penetration, as well as teaching tools and methods has evolved. Development this tools and methods in the sense that every person in every time and place, with facilities that can provided, that will determine the timeframe in which to engage in science learning. During the learning process and depending on events happening in the environment, learner's emotions are changed. In this situation, learning style should be revised according to the personality traits as well as the learner's current emotions. Virtual Learning, is one of the most frequently used terms associated with information technology has entered the educational field and Many educational institutions, especially universities, this part of the training programs have long term and do they mainly on investments in this category. Therefore, efforts and experiences related to this type of learning in the worldwide is highly regarded. In world, most universities are using this technology extensively. Some universities also accept students who take distance education. Virtual Education, a new field of communication technology and education which improve for learners, lifelong learning can provide at any time and place. In world, virtual education is widely considered. So with this kind of training will overcome many limitations of traditional education [1-8].

II. EDUCATION

In recent years, increasing demand for entry to university and study in any field is not hidden from anyone. Growing population of young professionals on the one hand and country needs for the proper design of industrial, agricultural and other areas on the other hand, will turn on given the need, new methods of training. Volunteers to respond to growing demand in the universities used the different strategies. So far, the quantity of academic development is the continued presence and Part-time. Development courses at night school, correspondence courses development, and participation by the private sector with the opening of foreign universities, including the way things are common. During recent years, use of virtual education, has been working in the universities program. This new technique is so promising, that even a young university, fully formed as a virtual. The University to that before the Web-based virtual training did not exist, now has several thousand Virtual students [9].

III. VIRTUAL EDUCATION

In the lexical, refers to all educational activities, using electronic tools, including audio, video, computer, network, and is virtual. In the conceptual, active learning and intelligent, the way in which developments in teaching and learning process and knowledge management, in develop and sustain cultural information and communication technology, the role will be pivotal. In fact, virtual education, distance education is based on technology [10]. Virtual Learning system emphasizes on the available content to all learners irrespective of their knowledge level and relevance. In other words, course content presented using voice and text files which using double relation between learner and teacher or among students, provide quality training to its highest reaches. Using advanced equipment and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

facilities to provide information and knowledge, better quality and higher provides [10, 11].

IV. Necessity, Importance and Objectives of Virtual Education

The growing needs of education, lack of access to education, lack of economic resources, lack of qualified educators, and many costs that are spent on education, the experts on that, with the help of information technology, new methods must be devised for both economic and quality and can be used to it, simultaneously a multitude of learners were trained. People want to continue college education has increased and with the current education system, only a few percent of the volunteers, they found an entry to the University. Given the recent developments and new global information age in which the highest value-added knowledge provides us with a major challenge has been met only with the benefits of virtual education can be overcome. The need for the development of virtual education in the country, there is no doubt; importance is the way how to achieve effective training. In general, the goal of virtual education, providing equal access, low cost and searchable in courses and creating uniform educational space for different classes of materials provided at any point and optimization methods for learning is deeper and more serious. In the educational environment unlike traditional education, those issues may take advantage of their ability [12].

V. Features of Virtual Education

Virtual Education has many features that can be the most important ones include [13]:

- *Complete mastery of the material*: Teachers in this way, always subject to question and criticize the competition with others, therefore, the issue of teacher training is not enough control, will not survive in the educational system.
- *Fair look to the knowledge seekers*: All segments of society to expand access to learning and opportunity, a great step forward for social justice in education.
- *Flexibility and tolerance*: In this manner, speed and talent of the courses offered is comprehensive and has changed and repeated discussions, there is no waste of time.
- *Audience Groups*: In the Virtual Education there are particular tools for audience group. Some of these tools include: assessment of candidates and determine the type of access set specific limits for each class of learners, the academic requirements to achieve some of the texts.
- *Free Education*: In learning there is a lot fields and conditions to closer to a free public education. Some of which include: reducing the cost of their education classes, no need to account for ancillary costs such as buildings, universities and etc.

VI. Groups that can useing of Virtual Education

Using Virtual Education, many groups can benefit from education. Some of these groups include [13]:

A. People living in remote areas

In many remote areas, people for various reasons, access to education for various reasons are not the person.

B. Women and girls

Gender differences in access to education, a very big challenge in developing countries, in these communities, is growing inequality between men and women (78 percent of the world's illiterate are women and girls.). Considering these issues, the need to educate girls and women and gender equality in access to education, the MDGs and the International Education for all was included.

C. People with physical defects

Virtual Education has provided the opportunity to help people overcome learning obstacles, obstacles such as printed materials, text, video and audio to the use of vision and hearing needs.

D. People outside the school

More than 130 million people worldwide do not have access to education. With the implementation of distance education, thousands people have been covered by the education system.

E. Workers and employees

In a world that is rapidly changing and transforming, lifelong learning is the only condition for survival and in fact, lifelong learning is a necessity for living in today's world. Hence, issues related to knowledge management and learning organization, each of the past are considered. Therefore, the work force to comply with new requirements and new technologies, they need to learn and learning, according to economic and time saving, it is the best source of training for employees.

VII. Comprehensive Skills Needed for Virtual Education

Skills that students need it for online learning, including interpersonal skills, study skills, general work skills with computers and the Internet [14].

A. Interpersonal skills

The nature of education at university level is changing. Increasingly, students are taking responsibility for learning. Students tend to give them all the questions teachers, should try to act as an active learner. The person responsible for learning, increasing motivation and discipline has its own, now more than ever has the opportunity to participate in learning, not just be a passive recipient. Students can make use the Internet to access a global community of students and teachers, therefore can be used of its benefits.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Study Skills

Although online education may be a new phenomenon for students, but some new cases, there are about education that despite the new technology, remain the same. Things like time management, motivation, expectations are clear and ready for the exam, are still as important aspects. Online education is related to reading and writing skills. Many of the thematic content of the reading material and offered an amount of correspondence, students will be in written form. If students do not have the ability to know that this relationship should be looking to develop these skills.

C. General Computer Skills

Students at the basic level of computer skills they need to succeed in online education act. Skills such as word processing, file management, storage, and publishing, although it is not necessary for students, would be helpful.

D. Internet Skills

Students for online study will need some skills to the Internet. Go to a specific address, search, save and print Web pages, are important skills. Advanced skills such as searching and evaluating Web site also will be useful for most students.

VIII. Defference Between Traditional Education and Virtual Education

In traditional education most attitudes is to skills and individual training. While in Virtual Education attitude is to social skills development of individuals. In traditional training, competitive spirit of the people make sweeping. Sometimes into the spirit of jealousy which has its own social consequences. While in Virtual Learning attention to context and environment interaction, one can simply create a spirit of partnership and teamwork in learning. It's a great source of research (Internet) that are readily available to learners and the possibility of any research group to provide for them. Because access to the Internet, content is also very flexible, so teachers can easily use it to keep its curriculum resource materials, while in traditional education, limited resources and has a few books and renewal and review of content, it might take years. Another point in Virtual Learning, using multimedia and simulation tools in the learning process. That allows learners to touch the virtual reality of what is supposed to learn [15]. While in traditional education, just with a few photographs or text or in the laboratory sessions, can be paid to training. Depending on the technology used, the type of attitude to class and the professor, as the main pillars of education will change. If the last class held a lecture by a professor or at best a question and answer, with Virtual Education, learning environment in a fully interactive environment that provides teachers and learners in this environment has become an observer and teaches a specific subject, but is a guide to self-learners. If we are in a traditional classroom in terms of location, time and cost constraints were held in virtual classrooms, there is no such restriction. Table.1 shows Differences between traditional teaching and Virtual Education [16].

Table.1 Traditional Teaching and Virtual Education

Dimensions	Traditional teaching	Virtual Education
Provides content	Coaching training imposes	Inclusive will choose the educational path
How to respond	Answers is pre-determined	Responses when faced with the problem
How progress and path learning	Pre-determined	Given the current situation and needs
Integration of learning with actions	Learning is a distinct activity with other works	Integrated discussion with other activities
Education and learning process	Format with specific start and end of the specified	Not stop and goes in parallel with business
Select materials and educational content	With selection teacher	Selection with interacting Inclusive and teacher
Compatible Content	In basic shape and Unchanged	Proportional with users and Flexible

IX. Improve the Quality of Teaching and Learning in Virtual Education

In the current era, the issue of education for all and lifelong learning is an accepted principle which negate the traditional look to the cross-training. One of the most fundamental reasons for using information and communication technologies in an educational system, is that the learning process for individual users and to facilitate curriculum. Allow to learners determine quickly to their learning and information resources are developed. Also, ICT can enhance active learning and interaction between learners and teachers in a flexible and constantly changing environment makes it possible to produce and distribute knowledge. Dynamic and challenging environment builds character, quality and increases effectiveness of learning [17]. Online learning environment at the university plays an important role in distance education can improve the quality of education [18]. Ways which through their internet learning environment can Improvement quality of education are:

- *Browse the Course*: Students can take courses offered through the Internet and they can read your speed.
- *Students will not ever lose your classroom*: Students in traditional education in addition to disease, possibly due to job obligations and family obligations, or of course they lose.
- *Traffic problems*: some students to attend class, should be over long distances and spend much time to traveling.
- *Easy access*: access to the information world that is achievable only through the World Wide Web. For

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

example, access to frequently asked questions, newsgroups, library catalogs and product information.

- *Increase Internet literacy*: literacy is a necessity in today's Internet, just as ten years ago computer literacy was a necessity.

X. EVALUATION OF COMPREHENSIVE IN VIRTUAL EDUCATION

One of the most worrisome in Virtual Education is assessing learners. Teachers are concerned which not able properly assessed level of understanding and participation of learners in the classroom. But evaluation of learners in Virtual Education is simpler than traditional education. In Virtual Education, learners can be record and store all responses. Results of examinations and assignments of students will be recorded in the memory device and used in the evaluation of them [18].

XI. INFRASTRAUCTURE NECESSITY FOR VIRTUAL EDUCATION

Virtual Education requires a lot of infrastructure that some of them are [2, 11]:

- Developing ICT skills at all levels of society to the public.
- Encourage and promote educational research in the field of information technology.
- Qualitative and quantitative expansion in the production of educational software.
- Equip schools and universities to computer and access to global network.
- Development of information education and communication skills.
- Strengthening the country's Internet network infrastructure.
- Level of public access to computers and networks worldwide.
- Development of IT in everyday culture.

XII. WAYS TO RESCUE EDUCATION FROM CRISIS

In the One of the main strategies for out of higher education in our country from current crisis, is according to E-Learning there is no doubt, but a simple look at the databases of universities that claim to their virtual education well implementation, this indicates that, related works were very preliminary and putting in a university course on site and is an e-mail boxes and other facilities limited which basically can not do this literally as virtual education and E-University [18]. In the virtual university's website, other than issues related to communication technologies, bandwidth and speed, reliable connection to the Internet (which is open to discussion), only a little better than the computer, certain categories of video programming and less of the characteristics of the virtual address can be observed and In fact, in distance education, computer technology to non-specialists and people familiar with the computer, what additional expertise should be used, major defects, is identifiable. For example, the unclear status of the educational technologist, curriculum planners, training and evaluation of curriculum, instructional designers and experts in teaching and learning strategies which traditional education in university and school were not used properly, will be how in Virtual Education and distance education university. One of the fundamental solution to out of the higher education from current crisis, according to eliminate the digital divide between our country and other countries and is also developing Virtual Education.

XIII. CONCLUSION

With the increasing spread of ICT and the public Internet will do many things outside of the traditional and new methods will be replaced. Education as one of the most basic needs will be no exception. In this context, Virtual Education can be as an excellent alternative to traditional education, but Virtual Education as a new way can be combined with learning and various teaching methods. Given the significant benefits of Virtual Education in comparison with traditional education and the progress of learners in E-Learning, obviously this method can be bring more satisfaction for students and faculty. Future prospects of virtual education would be Imagine which the free dissemination of knowledge between countries may lead to disputes between countries should be reduced. Given the proliferation of computer and Internet in training and advantages of virtual education in universities and educational system has increased in efficiency this system, universities can not ignore E-Learning. Hence the necessity of applying and implementing E-Learning systems to provide new services in teaching and learning has emerged as a fundamental requirement.

REFERENCES

[1] S. Fatahi, M. Ghasem-Aghaee, "Design and Implementation of an Intelligent Educational Model Based on Personality and Learner's Emotion", International Journal of Computer Science and Information Security, Vol. 7, No. 3, pp 1-13, 2010.

[2] M. Behrouzian-Nejad, E. Behrouzian-Nejad, "Electronic Education and Survey on infrastructure and challenges facing Electronic Education in Iran", Proceedings of First Regional Conference on new Achievement in Electrical and Computer Engineering, Islamic Azad University, Jouybar Branch, Iran, 2010.

[3] M. Behrouzian-Nejad, E. Behrouzian-Nejad, "effect of education and e-learning in improve the quality of teaching and learning and reduce costs", Proceedings of 13th Iranian Student Conference on Electrical Engineering, Tarbiat Modares Branch, Iran, 2010.

[4] S. Rafiei, S. Abdolahzadeh, "E-Learning in Medical Sciences", Scientific Research Center of Tehran University of Medical Sciences, Vol. 4, No. 13, 2009.

[5] A. R. Kamalian, A. Fazel, "Prerequisites and feasibility of implementing e-learning system", Journal of Educational Technology, Vol. 4, No. 1, 2009.

[6] M. A. Seyednaghavi, "Attitudes of teachers and students with e-learning: a survey of university education in Iran", Journal of Research and Planning in Higher Education, Vol. 13, No. 43, 2007.

[7] M. Behrouzian-Nejad, E. Behrouzian-Nejad, A. Ansariasl, "Survey on Barriers, constraints and infrastructure need to implementation of Virtual

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Education in Iran", Proceedings of 3th Iranian Conference on Electrical and Electronic Engineering, Islamic Azad University, Gonabad Branch, Iran, 2011.

[8] Medical Education Development Center, "Postgraduate medical education and training programs, to combine e-learning practices, through the Scholarships", 4th Festival Martyr Motahari, Shiraz, Iran, 2011.

[9] A. Kardan, A. Fahimifar, "Development of higher education, with an approach to virtual education: a response to needs, increase access and Challenges Ahead", Conference on Knowledge-based development, Iran, 2002.

[10] M. Behrouzian-Nejad, E. Behrouzian-Nejad, , "Survey on models and standards of Electronic Learning", Proceedings of 1th Regional Conference on Computer Engineering and Information Technology, Islamic Azad University, Dezfoul Branch, Iran, 2011.

[11] D. Venkatesan, RM. Chandrasekaran, "Adaptive e-Learning: A Conceptual Solution for the analysis of link between Medium of Instruction and Performance", International Journal of Computer Science and Information Security, Vol. 8, No. 5, pp 73-78, 2010.

[12] B. Niknia, "Necessary e-learning in today's world", Journal of Electronic Education, Vol. 15, No. 128, 2008.

[13] M. Atashak, "E-Learning: Concepts, findings and application", Proceeding of 3rd International Conference on Information Technology and knowledge, Ferdowsi University of Mashhad, Iran, 2007.

[14] H. Dargahi, M. Ghazisaeidi, M. Ghasemi, "Position of E-Learning in Medical Universities", Journal of School of Medicine Tehran University of Medical Sciences, Vol. 1, No. 2, 2007.

[15] R. Shafe'ei, "Anaglyph technology and its impact on the content, quality and attractiveness of education and learning", Proceedings of 2th National Conference on Information Technology Present, Feature, Islamic Azad University, Mashhad Branch, Iran, 2010.

[16] M. Yousefi, "E-learning needs of marine organisms in the near future", Journal of Marine Science, Imam Khomeini Noshahr, Special Section, No. 14, 2008.

[17] A. Ebrahimzadeh, H. Hasangholi, "Considerations in e-learning", Proceedings of Information Technology, http:\\ www.ahooeg.com, 08/04/2011.

[18] L. Molasalehi, R. Khalili, N. Jangjou, A. Khojastehband, A. Shahidi, A. Khalili, "Electronical University", Information Technology, Section 11, 2004.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Study of Neural Network Algorithm for Straight-Line Drawings of Planar Graphs

Mohamed A. El-Sayed[a], S. Abdel-Khalek[b], and Hanan H. Amin[c]

[a] Mathematics department, Faculty of Science, Fayoum University, 63514 Fayoum, Egypt
[b,c] Mathematics department, Faculty of Science, Sohag University, 82524 Sohag, Egypt
[a] CS department, Faculty of Computers and Information Science , Taif Univesity, 21974 Taif, KSA
[b] Mathematics department, Faculty of Science , Taif Univesity, 21974 Taif, KSA
[a] drmasayed@yahoo.com , [b] abotalb2010@yahoo.com, [c] hananhamed85@yahoo.com

***Abstract*— Graph drawing addresses the problem of finding a layout of a graph that satisfies given aesthetic and understandability objectives. The most important objective in graph drawing is minimization of the number of crossings in the drawing, as the aesthetics and readability of graph drawings depend on the number of edge crossings. VLSI layouts with fewer crossings are more easily realizable and consequently cheaper. A straight-line drawing of a planar graph G of n vertices is a drawing of G such that each edge is drawn as a straight-line segment without edge crossings.**

However, a problem with current graph layout methods which are capable of producing satisfactory results for a wide range of graphs is that they often put an extremely high demand on computational resources. This paper introduces a new layout method, which nicely draws internally convex of planar graph that consumes only little computational resources and does not need any heavy duty preprocessing. Here, we use two methods: The first is self organizing map known from unsupervised neural networks which is known as (SOM) and the second method is Inverse Self Organized Map (ISOM).

Keywords-SOM algorithm, convex graph drawing, straight-line drawing

I. INTRODUCTION

The drawing of graphs is widely recognized as a very important task in diverse fields of research and development. Examples include VLSI design, plant layout, software engineering and bioinformatics [13]. Large and complex graphs are natural ways of describing real world systems that involve interactions between objects: persons and/or organizations in social networks, articles incitation networks, web sites on the World Wide Web, proteins in regulatory networks, etc [23,10].

Graphs that can be drawn without edge crossings (i.e. planar graphs) have a natural advantage for visualization [12]. When we want to draw a graph to make the information contained in its structure easily accessible, it is highly desirable to have a drawing with as few edge crossings as possible.

A straight-line embedding of a plane graph G is a plane embedding of G in which edges are represented by straight-line segments joining their vertices, these straight line segments intersect only at a common vertex.

A straight-line drawing is called a convex drawing if every facial cycle is drawn as a convex polygon. Note that not all planar graphs admit a convex drawing. A straight-line drawing is called an inner-convex drawing if every inner facial cycle is drawn as a convex polygon.

A strictly convex drawing of a planar graph is a drawing with straight edges in which all faces, including the outer face, are strictly convex polygons, i. e., polygons whose interior angles are less than 180. [1]

However, a problem with current graph layout methods which are capable of producing satisfactory results for a wide range of graphs is that they often put an extremely high demand on computational resources [20].

One of the most popular drawing conventions is the straight-line drawing, where all the edges of a graph are drawn as straight-line segments. Every planar graph is known to have a planar straight-line drawing [8]. A straight-line drawing is called a convex drawing if every facial cycle is drawn as a convex polygon. Note that not all planar graphs admit a convex drawing. Tutte [25] gave a necessary and suifcient condition for a triconnected plane graph to admit a convex drawing. Thomassen [24] also gave a necessary and su.cient condition for a biconnected plane graph to admit a convex drawing. Based on Thomassen's result, Chiba et al. [6] presented a linear time algorithm for finding a convex drawing (if any) for a biconnected plane graph with a specified convex boundary. Tutte [25] also showed that every triconnected plane graph with a given boundary drawn as a convex polygon admits a convex drawing using the polygonal boundary. That is, when the vertices on the boundary are placed on a convex polygon, inner vertices can be placed on suitable positions so that each inner facial cycle forms a convex polygon.

In paper [15], it was proved that every triconnected plane graph admits an inner-convex drawing if its boundary is fixed with a star-shaped polygon P, i.e., a polygon P whose kernel (the set of all points from which all points in P are visible) is not

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

empty. Note that this is an extension of the classical result by Tutte [25] since any convex polygon is a star-shaped polygon. We also presented a linear time algorithm for computing an inner-convex drawing of a triconnected plane graph with a star-shaped boundary [15].

This paper introduces layout methods, which nicely draws internally convex of planar graph that consumes only little computational resources and does not need any heavy duty preprocessing. Unlike other declarative layout algorithms not even the costly repeated evaluation of an objective function is required. Here, we use two methods: The first is self organizing map known from unsupervised neural networks which is known as (SOM) and the second method is Inverse Self Organized map (ISOM).

II. PRELIMINARIES

Throughout the paper, a graph stands for a simple undirected graph unless stated otherwise. Let $G = (V,E)$ be a graph. The set of edges incident to a vertex $v \in V$ is denoted by $E(v)$. A vertex (respectively, a pair of vertices) in a connected graph is called a *cut vertex* (respectively, a *cut pair*) if its removal from G results in a disconnected graph. A connected graph is called *biconnected* (respectively, *triconnected*) if it is simple and has no cut vertex (respectively, no cut pair).

We say that a cut pair *{u, v} separates* two vertices *s* and *t* if *s* and *t* belong to different components in *G-{u, v}*.

A graph $G = (V,E)$ is called *planar* if its vertices and edges are drawn as points and curves in the plane so that no two curves intersect except at their endpoints, where no two vertices are drawn at the same point. In such a drawing, the plane is divided into several connected regions, each of which is called a *face*. A face is characterized by the cycle of G that surrounds the region. Such a cycle is called a *facial cycle*. A set F of facial cycles in a drawing is called an *embedding* of a planar graph G.

A *plane* graph $G = (V, E,F)$ is a planar graph $G = (V,E)$ with a fixed embedding F of G, where we always denote the outer facial cycle in F by $f_o \in F$. A vertex (respectively, an edge) in f_o is called an *outer vertex* (respectively, an *outer edge*), while a vertex (respectively, an edge) not in f_o is called an *inner vertex* (respectively, an *inner edge*).

The set of vertices, set of edges and set of facial cycles of a plane graph G may be denoted by $V(G)$, $E(G)$ and $F(G)$, respectively.

A biconnected plane graph G is called *internally triconnected* if, for any cut pair *{u, v}*, *u* and *v* are outer vertices and each component in *G - {u, v}* contains an outer vertex. Note that every inner vertex in an internally triconnected plane graph must be of degree at least 3.

A graph G is *connected* if for every pair $\{u, v\}$ of distinct vertices there is a path between *u* and *v*. The *connectivity* $\kappa(G)$ of a graph G is the minimum number of vertices whose removal results in a disconnected graph or a single-vertex graph K_1. We say that G is *k-connected* if $\kappa(G) \geq k$. In other words, a graph G is *3-connected* if for any two vertices in G are joined by three vertex-disjoint paths.

Define a plane graph G to be *internally 3-connected* if (a) G is 2-connected, and (b) if removing two vertices *u*,*v* disconnects G then *u*, *v* belong to the outer face and each connected component of $G-\{u, v\}$ has a vertex of the outer face. In other words, G is internally 3-connected if and only if it can be extended to a 3-connected graph by adding a vertex and connecting it to all vertices on the outer face. Let G be an n-vertex 3-connected plane graph with an edge $e(v_1,v_2)$ on the outer face.

III. PREVIOUS WORKS IN NEURAL NETWORKS

Artificial neural networks have quite long history. The story has started with the work of W. McCulloch and W. Pitts in 1943 [21]. Their paper presented the first artificial computing model after the discovery of the biological neuron cell in the early years of the twentieth century. The McCulloch-Pitts paper was followed by the publication from F. Rosenblatt in 1953, in which he focused on the mathematics of the new discipline [22]. His perceptron model was extended by two famous scientists in [2].

The year 1961 brought the description of competitive learning and learning matrix by K. Steinbruch [5]. He published the "winner-takes-all" rule, which is widely used also in modern systems. C. von der Malsburg wrote a paper about the biological self-organization with strong mathematical connections [19]. The most known scientist is T. Kohonen associative and correlation matrix memories, and – of course – self-organizing (feature) maps (SOFM or SOM) [16,17,18]. This neuron model has great impact on the whole spectrum of informatics: from the linguistic applications to the data mining.

The Kohonen's neuron model is commonly used in different classification applications, such as the unsupervised clustering of remotely sensed images.

In NN it is important to distinguish between supervised and unsupervised learning. Supervised learning requires an external "teacher" and enables a network to perform according to some predefined objective function. Unsupervised learning, on the other hand, does not require a teacher or a known objective function: The net has to discover the optimization criteria itself. For the unsupervised layout task at hand this means that we will not use an objective function prescribing the layout aesthetics. Instead we will let the net discover these criteria itself. The best-known NN models of unsupervised learning are Hebbian learning [14] and the models of competitive learning: The adaptive resonance theory [10], and the self-organizing map or Kohonen network which will be illustrated in the following section

The basic idea of competitive learning is that a number of units compete for being the "winner" for a given input signal. This winner is the unit to be adapted such that it responds even better to this signal. In a NN typically the unit with the highest response is selected as the winner[20].

M. Hagenbuchner, A.Sperduti and A.C.Tsoi described a novel concept on the processing of graph structured information using the self- organizing map framework which allows the processing of much more general types of graphs, e.g. cyclic

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

graphs [11] . The novel concept proposed in those paper, namely, by using the clusters formed in the state space of the self-organizing map to represent the ‘‘strengths’’ of the activation of the neighboring vertices. Such an approach resulted in reduced computational demand, and in allowing the processing of non-positional graphs.

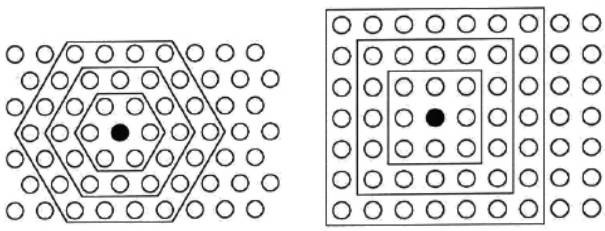
Georg PÄolzlbauer, Andreas Rauber, Michael Dittenbach presented two novel techniques that take the density of the data into account. Our methods define graphs resulting from nearest neighbor- and radius-based distance calculations in data space and show projections of these graph structures on the map. It can then be observed how relations between the data are preserved by the projection, yielding interesting insights into the topology of the mapping, and helping to identify outliers as well as dense regions [9].

Bernd Meyer introduced a new layout method that consumes only little computational resources and does not need any heavy duty preprocessing. Unlike other declarative layout algorithms not even the costly repeated evaluation of an objective function is required. The method presented is based on a competitive learning algorithm which is an extension of self-organization strategies known from unsupervised neural networks[20].

IV. SELF-ORGANIZING FEATURE MAPS ALGORITHM

Self-Organizing Feature Maps (SOFM or SOM) also known as Kohonen maps or topographic maps were first introduced by von der Malsburg [19] and in its present form by Kohonen [16].

According to Kohonen the idea of feature map formation can be stated as follows: The spatial location of an output neuron in the topographic map corresponds to a particular domain, or feature of the input data.



(a) Hexagonal grid (b) Rectangular grid

Figure 1. rectangular and hexagonal 2- dimensional grid

The general structure of SOM or the Kohonen neural network which consists of an input layer and an output layer. The output layer is formed of neurons located on a regular 1- or 2-dimensional grid. In the case of the 2- dimensional grid, the neurons of the map can exist in a rectangular or a hexagonal topology, implying 8-neighborhood or 6 neighborhoods, respectively. as shown in Figure (1).

The network structure is a single layer of output units without lateral connections and a layer of n input units. Each of the output units is connected to each input unit.

Kohonen’s learning procedure can be formulated as:

- Randomly present a stimulus vector x to the network
- Determine the "winning" output node u_i, where w_i is the weight vector connecting the inputs to output node i. $\|w_i - x\| \le \|w_j - x\| \forall k$

  Note: the above equation is equivalent to $w_i.x >= w_j.x$ only if the weights are normalized.
- Given the winning node i, and adapt the weights of w_k and all nodes in a neighborhood of a certain radius r, according to the function $w_i(new) = w_i(old) + \alpha.\Omega(u_i, u_j)(x - w_i)$
- After every j-th stimulus decrease the radius r and α.

Where α is adaption factor and $\Omega(u_i, u_j)$ is a neighborhood function whose value decreases with increasing topological distance between u_i and u_j .

The above rule drags the weight vector w_i and the weights of nearby units towards the input x.

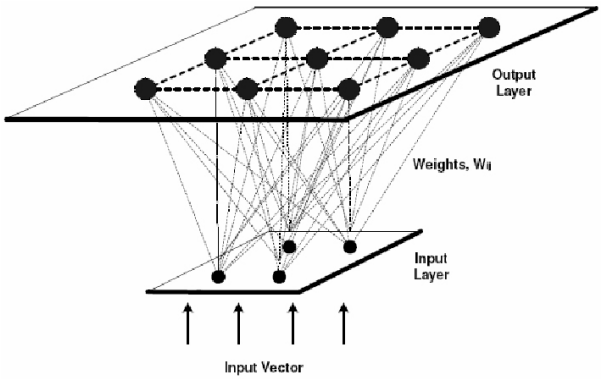


Figure 2. General structure of Kohonen neural network

This process is iterated until the learning rate α falls below a certain threshold. In fact, it is not necessary to compute the units’ responses at all in order to find the winner. As Kohonen shows, we can as well select the winner unit u_j to be the one with the smallest distance $\left\| \vec{v} - \vec{w}_j \right\|$ to the stimulus vector. In terms of Figure 3 this means that the weight vector of the winning unit is turned towards the current input vector.

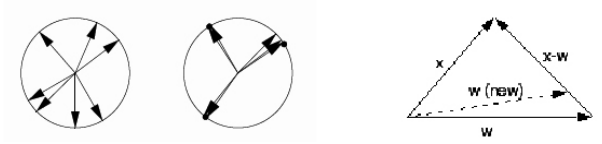


Figure 3. Adjusting the Weights.

Kohonen demonstrates impressively that for a suitable choice of the learning parameters the output network organizes itself as a topographic map of the input. Various forms are possible for these parameter functions, but negative exponential functions produce the best results, the intuition being that a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

coarse organization of the network is quickly achieved in early phases, whereas a localized fine organization is performed more slowly in later phases. Therefore common choices are: Gaussian neighborhood function $\Omega(u_i,u_j)=e^{-d(u_i,u_j)^2/2\sigma(t)^2}$ where $d(u_i,u_j)$ is the topological distance of u_i and u_j and σ^2 is the neighborhood width parameter that can gradually be decreased over time.

To get amore intuitive view of what is happening, we can now switch our attention to the weight space of the network. If we restrict the input to two dimensions, each weight vector can be interpreted as a position in two-dimensional space. Depicting the 4-neighborhood relation as straight lines between neighbors, Figure 4 illustrates the adaption process. Starting with the random distribution of weights on the left-hand side and using nine distinct random input stimuli at the positions marked by the black dots, the net will eventually settle into the organized topographic map on the right-hand side, where the units have moved to the positions of the input stimuli.

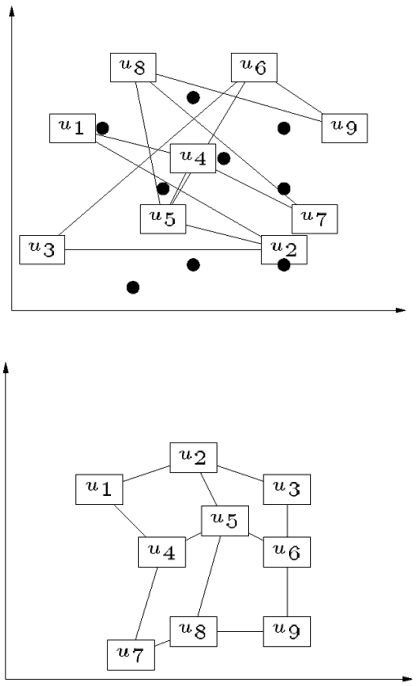


Figure 4. A Simple of random distribution of *G* and its the organized topographic map.

The SOM algorithm is controlled by two parameters: a factor α in the range 0…1, and a radius *r*, both of which decrease with time. We have found that the algorithm works well if the main loop is repeated 1,000,000 times. The algorithm begins with each node assigned to a random position. At each step of the algorithm, we choose a random point within the region that we want the network to cover (rectangle or hexagonal), and find the closest node (in terms of Euclidean distance) to that point. We then move that node towards the random point by the fraction α of the distance. We also move nearby nodes (those with conceptual distance within the radius *r*) by a lesser amount [11,20].

The above SOM algorithm can be written as the following:

input: An internally convex of planar graph *G*=(*V*,*E*)
output: Embedding of a planar graph *G*
radius *r* := r_{max}; /* initial radius */
initial learning rate α_{max} ;
final learning rate α_{min}
repeat many times
  choose random (*x*,*y*);
  i = index of closest node;
  move node *i* towards (*x*,*y*) by α ;
  move nodes with *d*<*r* towards (*x*,*y*) by $\alpha.e^{-d^2/2\sigma(t)^2}$.
  decrease α and *r*;
end repeat

V. INVERTING THE SOM ALGORITHM (ISOM)

We can now detail the ISOM algorithm. Apart from the different treatment of network topology and input stimuli closely resembles Kohonen's method [20].

In ISOM there are Input layer and weights layer only the actual network output layer is discarded completely in this method we look at the weight space instead of at the output response and to interpret the weight space as a set of positions in space.

The main differences to the original SOM are not so much to be sought in the actual process of computation as interpretation of input and output. First, the problem input given to our method is the network topology and not the set of stimuli. The stimuli themselves are no longer part of the problem description as SOM but a fixed part of the algorithm, we are not really using the input stimuli at all, but we are using a fixed uniform distribution. For this reason, the layout model presented here will be called the inverted self-organizing map (ISOM). Secondly, we are interpreting the weight space as the output parameter.

In this method, there is no activation function σ in difference of SOM. In ISOM we use a parameter called "cooling" (c) and we use different decay or neighboring function: In the SOM method we use the neighborhood function $\Omega(u_i,u_j)=e^{-d(u_i,u_j)^2/2\sigma(t)^2}$ where $d(u_i,u_j)$ is the topological distance of u_i and u_j and σ^2 is the width parameter that can gradually be decreased over time .

In ISOM we use the neighborhood function $\Omega(u_i,u_j)=-2^{-d(w_i,w_j)}$, where $d(w_i,w_j)$ is the distance between *w* and all successors w_i of *w*.

The above ISOM algorithm can be written as the following:

input: An internally convex of planar graph *G*=(*V*,*E*)
output: Embedding of a planar graph *G*
epoch *t* := 1;
radius *r* := r_{max}; /* initial radius */
initial learning rate α_{max} ;
cooling factor *c*;
forall *v* ∈*V* do *v.pos* := *random_ vector*();
while (*t* ≤ t_{max}) do

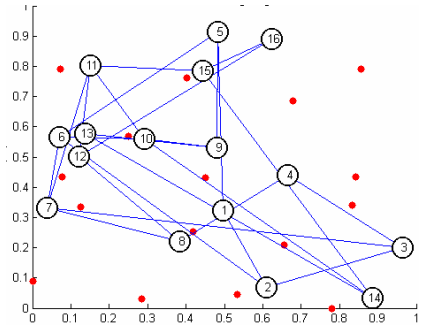
http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Adaption α := *max*(*min_adaption*,…
$e^{-c(t/t_{\max})}$.*max_adaption*)
i := *random_vector*();
/* uniformly distributed in input area */
w := *v* ∈*V* such that $\left\| v.pos - \vec{i} \right\|$ is minimal
for *w* and all successors w_i of *w* with $d(w,w_i) \le r$:
$w_i.pos = w_i.pos - 2^{-d(w_i,w_j)}.\alpha(w_i.pos - i)$;
t:=*t*+1;
if *r*> *min_radius* do *r*:=*r*-1;
end while.

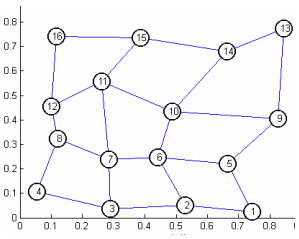
The node positions $w_i.pos$ which take the role of the weights in the SOM are given by vectors so that the corresponding operations are vector operations. Also note the presence of a few extra parameters such as the minimal and maximal adaption, the minimal and initial radius, the cooling factor, and the maximum number of iterations. Good values for these parameters have to be found experimentally [20].

VI. Experiments and Results

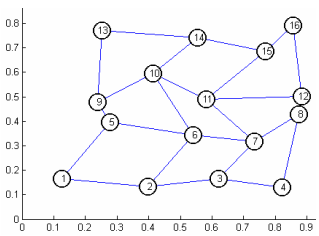
The sequential algorithm of the SOM model and ISOM were designed in Matlab language for tests. The program runs on the platform of a GIGABYTE desktop with Intel Pentium (R) Dual-core CPU 3GHZ, and 2 GB RAM.



(a) random weights of *G*



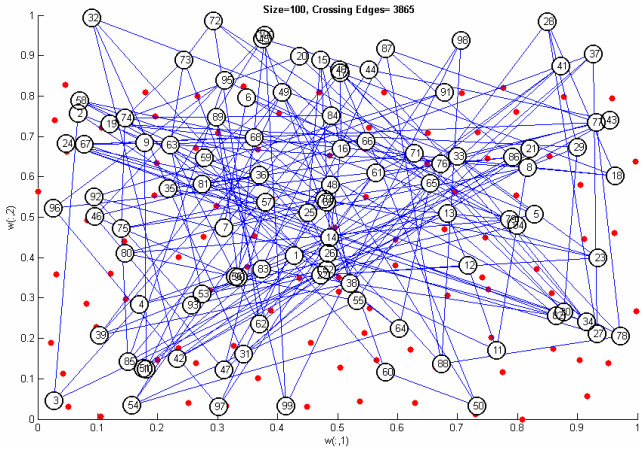
(b) SOM



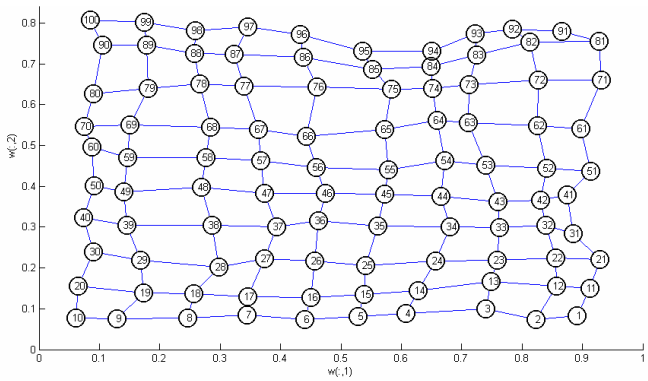
(c) ISOM

Figure 5. random weights of graph with 16 nodes, output graph drawing using SOM and ISOM, respectively.

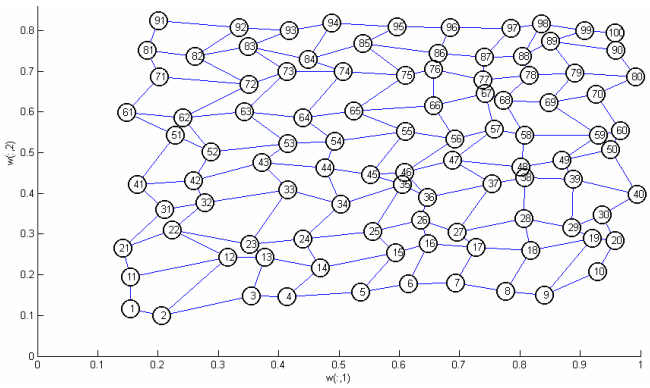
The algorithm was tested on randomly generated graphs *G*=(*V*,*E*). Initially, all vertices are randomly distributed in this area grid unit, and the weights generate at random distribution points .The initial graph has been drawing by many crossing edges see figure (5.a) where the grid size is (4*4) nodes.



(a) random weights of *G*, size=100 node , edge crossing = 3865



(b) SOM



(c) ISOM

Figure 6. random weights of graph with 100 nodes, output graph drawing using SOM and ISOM, respectively.

In the SOM method: The algorithm is controlled by two parameters: a factor α in the range 0…1, (we used initial

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

learning rate at α=0.5 and the final at α=0.1) and a radius r, (the initial radius at 3) both of which decrease with time.

In the ISOM method: The choice of parameters can be important. However the algorithm seems fairly robust against small parameter changes and the network usually quickly settles into one of a few stable configurations. As a rule of thumb for medium sized graphs, 1000 epochs with a cooling factor *c*=1.0 yield good results. The initial radius obviously depends on the size and connectivity of the graph and initial radius *r*=3 with an initial adaption of 0.8 was used for the examples in our paper. It is important that the intervals for radius and adaption both of which decrease with time. The final phase with *r*=0 should only use very small adaption factors (approximately below 0.15) and can in most cases be dropped altogether.

At each step of the algorithm, we choose random vector uniformly distributed in input area *i* and then find the closest node (in terms of Euclidean distance) between that point and the input stimuli. We then update the winner node and move their nearby nodes (those with conceptual distance within the radius *r*).

Each method generates a graph with minimum number of crossing, minimize the area of the graph and generate an internally convex planar graph. We have some examples as we can see in figures 5,6 .

We compare between three important isues: CPU time, drawing graph area in grid, and average length of edges using SOM and ISOM agorithms. In Table(1), The training time of the network effect directly on CPU time. So, we note that CPU time of SOM agorithm is less than ISOM agorithm. in compare with ISOM method. See the chart in figure 7.

TABLE I. CPU TIME,AREA,AND AVERAGE LENGTH OF EDGES

Example	Nodes of Graph	CPU time		Area		Average Length	
		SOM	*ISOM*	*SOM*	*ISOM*	*SOM*	*ISOM*
1	9	0.0842	0.0842	0.5072	0.3874	0.0752	0.0645
2	16	0.0936	0.0936	0.5964	0.5455	0.0397	0.0363
3	25	0.1310	0.1310	0.6102	0.5572	0.0212	0.0213
4	36	0.1498	0.1498	0.6438	0.6007	0.0142	0.0143
5	49	0.1872	0.1872	0.6479	0.6010	0.0103	0.0099
6	64	0.2278	0.2278	0.6800	0.6314	0.0077	0.0076
7	81	0.2465	0.2465	0.6816	0.6325	0.0060	0.0059
8	100	0.2870	0.2870	0.6677	0.6528	0.0049	0.0048
9	144	0.3962	0.3962	0.6983	0.6872	0.0034	0.0034
10	225	0.5710	0.5710	0.7152	0.6943	0.0021	0.0021

In VLSI applications, the small size of chip and the short length between the links are preferred. The main goals in our paper that minimize the area of output drawing graph on drawing grid, and minimize the average length of edges.

We note that ISOM method is better than SOM method to minimize the area and the average length of edges. In our experiments if the nodes greater than 400 nodes the SOM method generate graph with many crossing edges but ISOM generate graph no crossing edges in many times we train the program and ISOM is successes in minimize the graph area in compare with the SOM method .

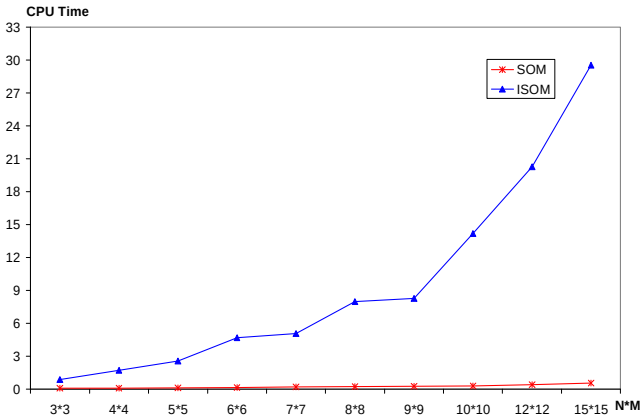


Figure 7. Chart of CPU time using SOM and ISOM, respectively

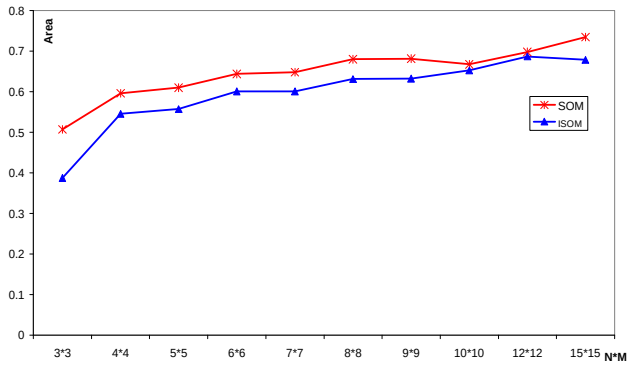


Figure 8. Chart of graph area using SOM and ISOM, respectively

VII. CONCLUSIONS

In this paper, we have presented two neural network methods (SOM and ISOM) for draw an *internally convex of planar graph*. These techniques can easily be implemented for 2-dimensional map lattices that consumes only little computational resources and don't need any heavy duty preprocessing. The main goals in our paper that minimize the area of output drawing graph on drawing grid, and minimize the average length of edges which can be used in VLSI applications, the small size of chip and the short. We were compared between them in three important issues: CPU time, drawing graph area in grid, and average length of edges. We were concluded that ISOM method is better than SOM method

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

to minimize the area and the average length of edges but SOM is better in minimize CPU time.

In future work we are planning to investigate three dimensional layout and more complex output spaces such as fisheye lenses and projections onto spherical surfaces like globes.

REFERENCES

[1] Imre Bárány and Günter Rote , "Strictly Convex Drawings of Planar Graphs", Documenta Mathematica 11, pp. 369–391, 2006.

[2] Arpad Barsi: Object Detection Using Neural Self-Organization. in Proceedings of the XXth ISPRS Congress, Istanbul, Turkey, July 2004.

[3] Eric Bonabeau a,b,, Florian Hhaux : Self-organizing maps for drawing large graphs, Information Processing Letters 67 , pp. 177-184, 1998.

[4] Lucas Brocki: Kohonen Self-Organizing Map for the Traveling Salesperson Problem, Polish–Japanese Institute of Information Technology, 2007.

[5] Carpenter, G.A., Neural network models for pattern recognition and associative memory. Neural Network, No. 2, pp. 243-257, 1989.

[6] N. Chiba, T. Yamanouchi and T. Nishizeki, Linear algorithms for convex drawings of planar graphs, Progress in Graph Theory, Academic Press, pp. 153-173, 1984.

[7] Anthony Dekker: Visualisation of Social Networks using CAVALIER , the Australian Symposium on Information Visualisation, Sydney, December 2001.

[8] I. F´ary, On straight line representations of planar graphs, Acta Sci. Math. Szeged, 11, pp. 229-233, 1948.

[9] Georg PÄolzlbauer, Andreas Rauber, Michael Dittenbach: Graph projection techniques for Self-Organizing Maps . ESSAN'2005 proceedings- European Symposium on Artifial Networks Burges(Belgium), pp. 27-29 April 2005, d-side publi, ISBN 2-930307-05-6

[10] S. Grossberg. "Competitive learning: from interactive activation to adaptive resonance." Cognitive Science, 11, pp. 23–63, 1987.

[11] M. Hagenbuchner, A.Sperduti, A.C.Tsoi: Graph self-organizing maps for cyclic and unbounded graphs, Neurocomputing 72, pp. 1419–1430, 2009

[12] Hongmei. He, Ondrej. Sykora: A Hopfield Neural Network Model for the Outerplanar Drawing Problem, IAENG International Journal of Computer Science, 32:4, IJCS_32_4_17 (Advance online publication: 12 November 2006)

[13] Seok-Hee Hong and Hiroshi Nagamochi : Convex drawings of hierarchical planar graphs and clustered planar graphs, Journal of Discrete Algorithms 8, pp. 282–295, 2010.

[14] J. Hertz, A. Krogh, and R. Palmer. Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City/CA, 1991.

[15] S.-H. Hong and H. Nagamochi, Convex drawings with non-convex boundary, 32nd International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2006) Bergen, Norway June 22-24, 2006.

[16] T. Kohonen, , Correlation matrix memories. IEEE Transactions on Computers, Vol. 21, pp. 353-359, 1972.

[17] T. Kohonen, , Self-organization and associative memory. Springer, Berlin, 1984.

[18] T. Kohonen, , Self-organizing maps. Springer, Berlin, 2001.

[19] Malsburg, C. von der, Self-organization of orientation sensitive cells in the striate cortex. Kybernetik, No. 14, pp. 85-100, 1973.

[20] Bernd Meyer: Competitive Learning of Network Diagram Layout. Proc. Graph Drawing '98, Montreal, Canada, pp. 246–262, Springer Verlag LNCS 1547.S.

[21] R. Rojas, , Theorie der neuronalen Netze. Eine systematische Einführung. Springer, Berlin,1993.

[22] F. Rosenblatt, , The perception. A probabilistic model for information storage and organization in the brain. Psychological Review, Vol. 65, pp. 386-408, 1958.

[23] Fabrice Rossi and Nathalie Villa-Vialaneix: Optimizing an organized modularity measure for topographic graph clustering: A deterministic annealing approach , Preprint submitted to Neurocomputing October 26, 2009

[24] C. Thomassen, Plane representations of graphs, in Progress in Graph Theory, J. A. Bondy and U. S. R. Murty (Eds.), Academic Press, pp. 43-69, 1984.

[25] W. T. Tutte, Convex representations of graphs, Proc. of London Math. Soc., 10, no. 3, pp. 304-320, 1960.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Multithreaded Image Processing

Jamil A. M. Saif
Computer Science Department
Faculty of Computer Science and engineering
Hodeidah University
Hodeidah, Yemen
e-mail: jamil_alabssi@yahoo.com

Hamid S. S. Alraimi
Computer Science Department
Faculty of Computer Science and engineering
Hodeidah University
Hodeidah, Yemen
e-mail: halraimi@gmail.com

***Abstract*— real time image processing applications require a huge amount of processing power, computing ability and large resources to perform the image processing applications. The nature of processing in typical image processing algorithms ranges from large arithmetic operations to fewer one.**
This paper presents an implementation of image processing operations using simultaneous multithreading, the performance of multithreading is analyzed and discussed, for the varying number of images.

Keywords- multithreading; image processing; performance.

I. INTRODUCTION

Recently digital image processing has a broad spectrum of applications, such as multimedia systems, business systems, monitoring, inspection systems, and archiving systems. In spite of digitization, storage, transmission, and display operations, extra functions are considered. They are as follows: image data compression and representation, image enhancement and reconstruction, image indexing, retrieval and matching, etc. and they are executed on application oriented servers.

Generally three levels of image processing are distinguished to analyze and tackle the image processing application[1]: low-level operations, intermediate-level operations, and high-level operations.

Low-level operations: Images are transformed into modified images. These operations Work on whole image structures and yield an image, a vector, or a single value. The computations have a local nature; they work on single pixels in an image. Examples of Low-level operations are: smoothing, convolution and histogram generation.

An intermediate-level operations: Images are transformed into other data structures. These operations work on images and produce more compact data structures (e.g. a list). The computations usually do not work on a whole image but only on objects/segments (so called regions of interest ROI) in the image. Examples of intermediate-level operations are: region labeling and motion analysis.

A high-level operations: Information derived from images is transformed into results or actions. These operations work on data structures (e.g. a list) and lead to decisions in the application. So high-level operations can be characterized as symbolic processing. An example of a high-level operation is object recognition.

There is a big challenge concerning image processing due to time consuming computation, some researches address this problem using parallel environments[2,5] such as PVM, MPI, others used distributed parallel processing using java RMI, Sockets and Corba[4].

In image processing operations the existing approach to parallelism get constrained due to variant size of data and the required resources. Hence a system is required for the efficient controlling of image processing operation with variable data size. for this reason a multithreading approach is proposed.

The contents of this paper is organize as follows :in section 2 image conversion is presented, in section 3 a multithreading and its related concepts are defined, in section 4 the results obtained from the experiments are described and discussed, finally the summarized conclusion is given.

II. IMAGE CONVERSION

In this paper a low level image processing is used that will modify RGB colored image into grey scale one, the RGB image is transformed according to the following formula [6]:

$$I = \alpha_1 R + \alpha_2 G + \alpha_3 B \quad \text{.....(1)},$$

$$\text{where : } \alpha_1 + \alpha_2 + \alpha_3 = 1, \text{and I is grey scale value}$$

For each pixel in RGB image the I grey scale value is calculated and this calculation is repeated by scanning the whole image starting from the upper left corner to the bottom right corner of the given image, and this calculation may be required for several images, these heavy computations need some way to reduce the cost of computation.

III. MULTITHREADING AND ITS RELATED CONCEPTS

Multithreading is a technique that allows a program or a process to do many tasks concurrently at the same time [9,10]. Multithreading allows a process to run tasks in parallel on a symmetric multiprocessing (SMP) system or a chip multithreading [7,8] (CMT) system, allowing the process to scale linearly with the number of cores or processors, which improves performance, increases efficiency, and increases throughput.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Running multiple processes concurrently is called multiprocess programming. A process is a heavyweight entity that lives inside the kernel. It consists of the address space, registers, stack, data, memory maps, virtual memory, user IDs, file descriptors, and kernel states. Where as a thread is a lightweight entity that can live in the user space or the kernel and consists of registers, stack, and data. Multiple threads share a process, that is, they share the address space, user IDs, virtual memory, file descriptors, and kernel states. The threads within a process share data, and they can see each other, to distinguish between a process and a thread see Fig. 1, where two threads within one process.

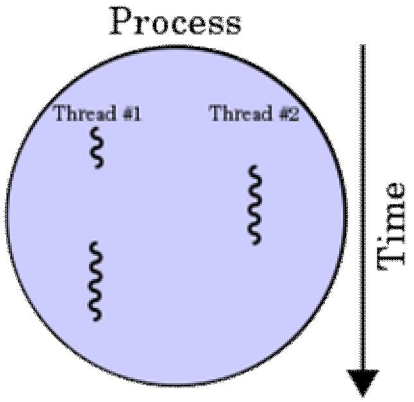


Figure 1. A process with two threads of execution.

Multithreading [8] is a way of achieving multitasking in a program. Multitasking is the ability to execute more than one task at the same time see Fig. 2. Multitasking can be divided into Process-based multitasking and thread-based multitasking.

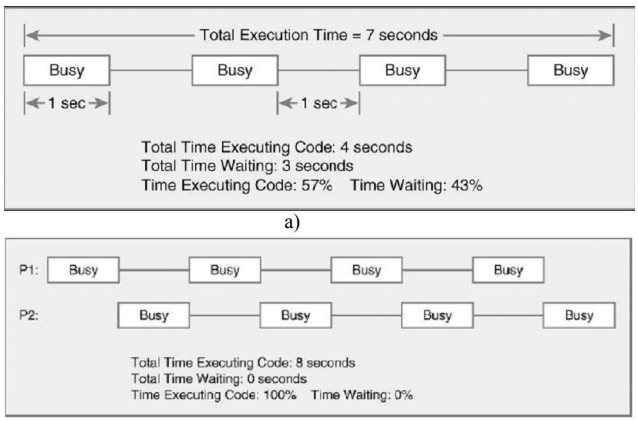


Figure 2. a) with single task b) with two tasks

Process-based multitasking feature enables you to switch from one program to another so fast that it appears as if the programs are executing at the same time. Where as thread-based multitasking context-switch is extremely fast and can be in user space or at the kernel central processing unit (CPU) level. A process is heavyweight, so it costs more to context-switch than a thread.

A single program can contain two or more threads and therefore, perform two or more tasks simultaneously see Figure 2. A text editor can perform writing to a file and print a document simultaneously with separate threads performing the writing and printing actions. In the text editor, you can format text in a document and print the document at the same time. There are fewer overloads when the processor switches from one thread to another. Therefore, threads are called lightweight process. On the other hand, when the processor switches from one process to another process the overload increases. Advantages of multithreading are: improved performance, minimized system resource usage, simultaneous access to multiple applications and program structure simplification. Improved performance provides improvement in the performance of the processor by simultaneous execution of computation and the I/O operation see Fig 2 . . Minimized system resource usage minimizes the use of system resources by using threads, which are the same address space and belong to the same process. Simultaneous access to multiple applications provides access to multiple applications at the same time because of quick context switching among threads. A thread is lightweight, so many threads can be created to use resources efficiently. The threads are all within a process see Figure 1, so they can share global data. A blocking request by one thread will not stop another thread from executing its task. Also, the process will not get context-switched because a thread is blocked.

Multiprocess programming is much more difficult than multithreaded programming, performance is slower, and management of resources is difficult. Also, synchronization and shared memory use are more difficult with processes than with threads, because threads share memory at the process level and global memory access is easy with threads.

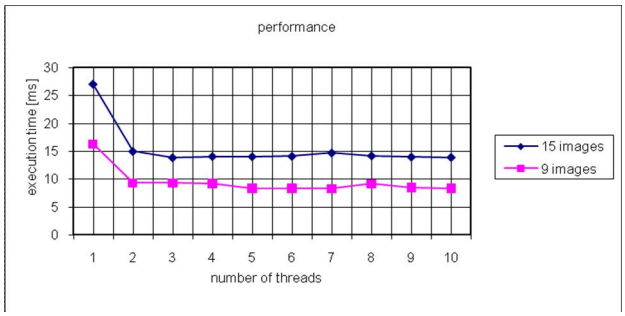
The result of multithreading is increased performance, increased throughput, increased responsiveness, the ability to execute tasks repeatedly, increased efficiency, better management of resources, and lowered costs [3,7].

IV. EXPERIMENTS

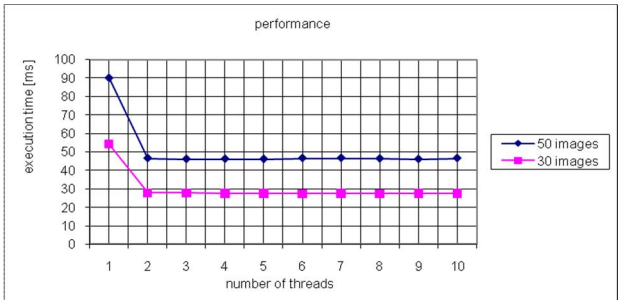
.Net environments for implementing multithreading image conversion were used, testing the multithreading with variable number of RGB colored images (9, 15, 30 and 50) each of 600x400 pixels of size, converting images into grey scale according to the formula 1., that image conversion carried out using single thread as well as multithreading varies from 2 to 10 threads.

The obtained results are shown in Figure 3.,demonstrate the efficiency of multithreading. As noticed every image took around 2 [ms] of computation, and since for our experiments a laptop with a dual core cpu of 3000 MHz was used, at least two threads are needed to fully utilize the two cores, so that is illustrated in the Fig. 3, using two threads cause reducing execution time to about 50% , while for three threads and more some slight improvement is seen, and as the date size increase (number of images) the performance almost remains the same, this is due to the multithreading overhead in comparison with computation time.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



a)



b)

Figure 3. a) multithreading for 9 and 15 images
b) multithreading for 30 and 50 images

V. CONCLUSION AND RECOMMENDATION

Image processing is a time consuming computation, for improving performance, multithreading was used. It is obvious from the Figure 3 the impact of date size and the contributing threads on the performance.

It is recommended to use in future work heavier computation that needs significant time to enable showing the advantage of threads addition and utilize environment with more cores or processors to demonstrate the scalability of such systems.

REFERENCES

[1] K. J. Anil, "Fundamentals of digital image processing", Prentice Hall, april, 2004.
[2] P. Czarnul, H. Krawczyk, "Dynamic Assignment with Process Migration in Distributed Environments. Recent Advances in PVM and MPI", Lecture Notes in Computer Science, Vol. 1697, 1999.
[3] O. Edelstein, E. Farchi, Y. Nir, G. Ratsaby and S. Ur, "Multithreaded Java Program Test Generation", IBM SYSTEMS JOURNAL, VOL 41, NO 1, 2002.
[4] R. Eggen and M. Eggen, "Efficiency of Distributed Parallel Processing Using Java RMI, Sockets and Corba, 2007.
[5] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek and V. Sunderam, "PVM Parallel Virtual Machine. A User Guide and Tutorial for Networked Parllel Computing.", Mit Presso, Cambridge, 1994,http://www.epm.ornl.gov/pvm.
[6] W. Malina, S. Ablameyko and W. Pawlak, "Fundamental Methods of Digital Image Processing. (In Polish), 2002.
[7] J. Manson and W. Pugh, "Semantics of Multithreaded Java", 2002.
[8] N. Padua-Perez and B. Pugh, "multithreading in Java".
[9] H. Schild, "C# The Complete Reference", Edition McGraw-Hill, 2002.
[10] H. Shildt, "Java 2 The Complete Reference", Fifth Edition McGraw-Hill, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A New Efficient Symbol Timing Synchronization Scheme for MB-OFDM UWB Systems

Reza Shahbazian
Department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
Shahbazian@elec.iust.ac.ir

Bahman Abolhassani
Department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
Abolhassani@iust.ac.ir

***Abstract*— Conventional symbol timing synchronization algorithms show improper performance in low SNR values. In this paper a new low complexity and efficient symbol timing synchronization (ESTS) algorithm is proposed for MB-OFDM UWB systems. The proposed algorithm locates the start of Fast Fourier Transform (FFT) window during packet/frame synchronization (PS/FS) sequences of the received signal. First, a cross correlation based function is defined to determine the time instant of the useful and successfully detected OFDM symbol. The threshold value in detection of the OFDM symbol is predetermined by considering the trade-off between the probability of false alarming and missed detection. The exact boundary of the FFT window for each OFDM symbol is estimated by a maximum likelihood metric and choosing the argument of the peak value. Verifying the estimated timing offset is the last step to locate the start of the FFT window. The proposed algorithm shows great improvement in the MSE, synchronization probability and bit error rate metrics compared with those of earlier works.**

Keywords- MB-OFDM, Synchronization, Ultra Wide Band, Fast Fourier Transform, Maximum Likelihood.

I. INTRODUCTION *(HEADING 1)*

Ultra-Wideband (UWB) technology is the main candidate for short distance (<10 m) and high data rate (53-480 Mbps) communications in Wireless Personal Area Networks (WPAN). Multi band orthogonal frequency division multiplexing (MB-OFDM) based communication scheme is the most noteworthy, among the several proposals for efficient use of the 7.5 GHz bandwidth allocated for UWB technology.

MB-OFDM is the combination of OFDM modulation and data transmission using frequency-hopping techniques. In this method, all the available bandwidth (3.1-10.6 GHz) is divided into 14 frequency bands each with 528 MHz of bandwidth. These 14 frequency bands are categorized in five groups. Each of the first four groups has three frequency bands and the fifth group contains only two frequency bands. Data is transmitted over different frequency bands using a Time-Frequency code (TFC), which causes frequency diversity and multiple access capability [1].

OFDM systems have the advantage of being able to operate as a set of N (number of subcarriers in the system) parallel links over flat fading channels. However, the performance of non-ideal OFDM systems is degraded by imperfections caused by timing offset, improper number of cyclic prefix (CP) and frequency offsets. Among all the imperfections, effect of timing offset on the system performance and bit error rate is much more sever. Synchronization techniques for narrowband OFDM systems utilize maximum correlation between the received signal and training timing symbols [2-3]. All such techniques assume that the first received multipath component (MPC) is the strongest one. Therefore, in a channel with dense multipath effects, a delayed stronger component, which is shown in "Fig 1", may cause erroneous timing synchronization, which leads to Inter Symbol Interference (ISI), destroys the orthogonality of OFDM subcarriers, and degrades the overall performance [4].

Several algorithms are proposed for timing synchronization in MB-OFDM systems [5-9]. In [5], the proposed algorithm (FTA) detects the significant path by comparing the difference between two consecutive accumulated energy samples at the receiver against a predetermined threshold. However, the threshold is only determined by the probability of false alarm, while other important error measures such as the missed detection probability is not exploited. Further, the computational complexity is high due to the large amount of multiplications involved in the algorithm. In [6], a correlation based symbol timing synchronization (CBTS) has also been reported. The idea is similar to that of [5] and estimates the first significant multipath of the received signal by comparing the difference between two successive correlated MB-OFDM symbols against a predetermined threshold. Compared with that of [5], the computational complexity is reduced and performances in terms of both the mean square error (MSE) of timing offset and the perfect synchronization probability are improved. These two algorithms [5-6] cannot operate properly at low SNR values due to imperfections in autocorrelation property of the base sequence and the dense multipath channel environments. Combination of the autocorrelation function and restricted and normalized differential cross-correlation (RNDC) with a threshold-based detection is used in [7] to find the timing offset of the OFDM symbol. In [8], the proposed algorithm utilizes a maximum likelihood function to estimate the timing offset. Concentration of the algorithm in [8] is on frequency diversity. Moreover its computational complexity is rather high. In this paper, a modified and Efficient Symbol Timing Synchronization (ESTS) algorithm for MB-OFDM

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

UWB systems is proposed, while utilizes time domain sequences (TDS) to estimate the timing offset. The computational complexity of the proposed algorithm is reduced by simplification in correlation based and maximum likelihood functions. The organization of this paper is as follows: in Section II, we present the MB-OFDM system, signal model and characteristics of an UWB channel. In Section III, we describe the proposed algorithm for MB-OFDM timing synchronization and Section IV shows simulation results of our proposed algorithm and compares them with those reported in [5-9]. Important concluding remarks are made in Section V.

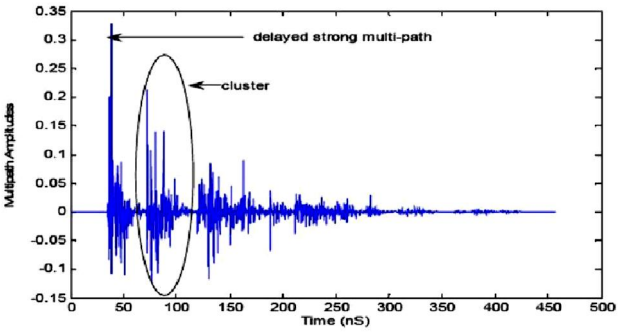


Figure 1. Impulse response of an UWB channel [9]

II. MB-OFDM System Model

A. MB-OFDM Signal Model

Synchronization in MB-OFDM systems is data-aided [1]. In standard preamble structure, the first 21 packet synchronization (PS) sequences are used for packet detection, AGC stabilization, coarse timing and frequency synchronization. The next 3 frame synchronization (FS) sequences are meant for a fine timing and frequency synchronization.

These sequences are followed by 6 channel estimation (CE) sequences as shown in "Fig 2". Depending on the time-frequency code, a particular preamble pattern is selected which is shown in "Table 1". For a given TFC the PS and FS sequences have the same magnitude but opposite polarity. The preamble structure for TFC 1 and 2 is shown in "Fig 2". Delay period is defined as the minimum number of symbol timing difference in the same frequency band. As an illustration, the delay period=3 for TFC 1 or 2, delay period=6 for TFC 3 or TFC 4 patterns and delay period=1 for TFC 5.

Consider $S_{s,n}(k)$ as k[th] sample of n[th] transmitted OFDM symbol, which is given by.

$$S_{s,n}(k) = S_c(n) \times S_b(k). \quad (1)$$

In "(1)", $S_b(k)$ is the k[th] sample of the nth symbol [11]. $S_b(k)$ is a time domain base sequence that is chosen according to the TFC employed and $S_c(n)$ is the spreading sequence for the nth symbol and $k = 1,2,...,M$ and $n = 1,2,...,P$, which M is the number of useful samples in one OFDM symbol and P is the total number of transmitted symbols in PS, FS and CE sequences. MB-OFDM symbols prepared by suffixing 32 null samples called zero padded (M_{ZP}) and 5 null guard samples called (M_g) to FFT/IFFT output sequences of length M which is considered to be 128 samples according to the frame format [11]. The total length of $M+M_{zp}+M_g$ samples of one MB-OFDM symbol is denoted by MT, which is equal to 165 samples.

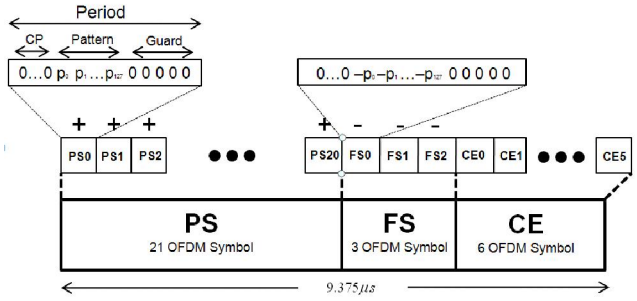


Figure 2. Packet model for a MB-OFDM system [1]

TABLE I. TFC PATTERN IN MB-OFDM SYSTEMS [1]

TFC Number	Preamble Number	TFC					
1	1	1	2	3	1	2	3
2	2	1	3	2	1	3	2
3	3	1	1	2	2	3	3
4	4	1	1	3	3	2	2
5	5	1	1	1	1	1	1
6	5	2	2	2	2	2	2
7	5	3	3	3	3	3	3

B. UWB Channel Model

IEEE802.15.3 channel modeling sub-committee has specified 4 different channel models (CM1-CM4) depending on transmission distances based on a modified saleh-valenzuela (S-V) model [10]. UWB channel model is a cluster-based model, where individual ray shows independent fading characteristics. An UWB channel not only shows frequency dependence of instantaneous channel transfer functions, but also the variations of averaged transfer function caused by different attenuations of different frequency component of an UWB signal [12].

Impulse response model of an UWB channel can be represented as,

$$h(t) = \sum_{l=0}^{L} \sum_{k=0}^{K} a_{k,l} \ \exp(j\varphi_{k,l}) \ \delta(t - T_l - \tau_{k,l}). \quad (2)$$

In "(2)", $\{a_{k,l}\}$ and $\{\varphi_{k,l}\}$ are tap weighting coefficients and tap phases of the k^{th} component in l[th] cluster respectively, and $h(t)$ represents small scale fading amplitude. Delay of k^{th} MPC toward arrival time of l[th] cluster, $\{T_l\}$, is shown with $\{\tau_{k,l}\}$. We denote $\mathbf{h}(t) = [h(0), h(1), ..., h(L-1)]$ as the channel

Identify applicable sponsor/s here. *(sponsors)*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

impulse response with L resolvable multipath components. We also define $n(t)$ as a zero mean additive white Gaussian noise (AWGN) with variance σ_n^2. The received signal with timing offset equal to θ could be described as the following,

$$r(k) = \sum_{i=0}^{L-1} S_s(k-\theta).h(i) + n(k). \quad (3)$$

III. PROPOSED ESTS ALGORITHM

The main objective in the symbol timing synchronization is to find the timing offset of the received symbol. Our proposed algorithm contains two steps, coarse and fine synchronization. The aim in coarse synchronization is to determine the time instant while a useful OFDM symbol has been successfully detected. In fine synchronization, we use the gained boundary in coarse synchronization to locate the exact starting point of the FFT window. To do synchronization, we use modified cross correlation based functions, which perform better than auto correlation functions, in low SNR values. The cross correlation function in general could be defined as,

$$F'_P(\theta) = \sum_{k=0}^{M-1} r(k+\theta).S^*_b(k). \quad (4)$$

In "(4)", the operator $(.)^*$ represents complex conjugate transpose of the signal and F_P indicates the crosscorrelation function between the received signal and the base sequence. The estimated and coarse boundary for timing offset could be found by the following Maximum Likelihood metric,

$$\hat{\theta} = \arg\max_{\theta} \{\psi(\theta)\} = \arg\max_{\theta} \left\{ \frac{|F'_P|^2}{|F_R|^2} \right\}, \quad (5)$$

Where we define

$$F_R(\theta) = \frac{1}{2}\left(\sum_{k=0}^{M-1} |r(k+\theta)|^2 + \sum_{k=0}^{M-1} |S_b(k)|^2 \right). \quad (6)$$

Computational complexity in this method is high. As shown in [13] we can use a simplified timing metric, which is a good approximation of "(5)", described as,

$$\lambda_F(\theta) = |\mathrm{Re}(F'_p(\theta))| + |\mathrm{Im}(F'_p(\theta))|. \quad (7)$$

If the base sequence is characterized by a perfect autocorrelation property, there is only one significant peak located at the first received sample. However, by imperfect autocorrelation property of the base sequence, as indicated in [1], there exist some undesired peaks at the other sample instants. By considering the AWGN and channel variations, these undesired peaks may be amplified and their values are comparable with that of the first peak corresponding to the desired symbol boundary. So the crosscorrelation function may trigger false alarm and the algorithms, which use these kinds of functions [5-9] show poor system performances. In order to reduce false alarm probability especially at low SNR values, we modify the introduced metric in "(7)" as the following,

$$\lambda(\theta) = |\mathrm{Re}(F_p(\theta)) \times \mathrm{Im}(F_p(\theta))|. \quad (8)$$

The defined function performs well at all SNR values if it is assumed that the packet is successfully detected and the OFDM sequences are confirmed to be received. In practical scenarios there exists a noise sequence at the start of every frame [14] which makes us to do a kind of packet detection at the start of timing synchronization algorithm but Computational complexity is rather high and needs M multiplications just in one crosscorrelation function. So, we reduce the complexity by simplifying "(4)" as described below,

$$F_P(\theta) = \sum_{k=0}^{M-1} r(k+\theta).\mathrm{sgn}(S_b(k)). \quad (9)$$

Define $V_{m+1} = [\theta+m, 1, ..., \theta+m+M-1]$ as the time index that contains the sign of $(m+1)^{th}$, M sample base sequence. Also, define $V_0 = [0,1,...,M-1]$ as the time index that contains the sign of M sample base sequence. We use M instead of M_T because there is no useful information in M_{ZP} and M_g sequences, i.e., $S_b(k) = 0 \quad M < k < M_T$. We assume that the channel and the noise are uncorrelated. The cross correlation function at time instant $\theta+m+k$ is given by:

$$E\left\{ \sum_{k=0}^{M-1} r(\theta+m+k).\mathrm{sgn}(S_b(k)) \right\} \quad (10)$$

This can be easily shown that by expanding "(10)" we can drive the following formula,

$$\sum_{k=0}^{M-1-m} S_c(n).|S_b(m+k)|.\mathrm{sgn}(S_b(m+k)).\mathrm{sgn}(S_b(k)).E\left\{ \sum_{k=0}^{L-1} h(k) \right\}. \quad (11)$$

In "(11)", when $m=0$, a negative and positive peak of the crosscorrelation is generated if $S_c(n) = -1$ and $S_c(n) = +1$ respectively. It means that when the time index that contains the first M sample of the received signal is considered, the peak value is generated. So, we use two sets of V_0 and V_1 for symbol timing offset estimation.

As the timing offset decreases the value of $\lambda(\theta)$ in "(8)" increases. We define S_N as the index of a received M sample sequence and $\omega(S_N)$ as the time instant of the first sample for that sequence.

$$\omega(S_N) = \arg_{\omega} \{ \lambda(\omega(S_N)) \geq \xi \}, \quad (12)$$

where $\lambda(\omega(S_N)) = |\mathrm{Re}(F_p(\omega(S_N))) \times \mathrm{Im}(F_p(\omega(S_N)))|$ and F_P is defined in "(9)". Parameter ξ in "(12)" is the threshold which is predetermined by considering the trade-off between the probability of false alarming and the probability of missed detection. If the OFDM symbol is successfully detected the value $\omega(S_N)$ is used as a reference symbol boundary for fine synchronization. Due to the modified S-V channel model, the first arriving path may not be the strongest one. As a result, using only the conventional cross-correlation function will

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

locate a delayed multipath component with stronger amplitude as the reference one and hence will cause misdetection. To correctly estimate the position of the first arriving path, we take the moving average of $\lambda(\omega(S_N))$ over a window of size L' where most of the channel energy is concentrated. In other words,

$$\lambda^{'}(\omega(S_N)) = \sum_{w=0}^{L'-1} \lambda(\omega(S_N)+w). \quad (13)$$

To reduce the computational complexity the "(13)" could be substituted by the following recursive equation as given below,

$$\lambda^{'}(\omega(S_N)+1) = \lambda^{'}(\omega(S_N)) + \lambda(\omega(S_N)+L') - \lambda(\omega(S_N)). \quad (14)$$

In "(14)", L' is considered as the maximum delay spread of the multipath channel. The exact symbol boundary ($\omega^o(S_N)$) could be found by the following equation

$$\omega^o(S_N) = \arg\max_{\omega} \left\{ \lambda^{'}(\omega(S_N)), \lambda^{'}(\omega(S_N)+1), ..., \lambda^{'}(\omega(S_N)+M-1) \right\}. \quad (15)$$

If the calculated value of $\omega^o(S_N)$ in "(15)", stands in the range of added zero prefix (M_{ZP}), all the subcarriers would experience the same phase shift that could be removed in the receiver. And if the $\omega^o(S_N)$ value stands out of this range, ISI occurs and subcarriers try different phase shifts that degrade the system performance. Since transmission channel varies in time, timing offset of each symbol is different from the others. Detailed flowchart of the proposed algorithm (ESTS) is shown in "Fig 3". When the estimated value stands in the ISI free zone (sample index $1 \rightarrow M_{ZP}$), synchronization is done. If the estimated value stands in the sample index $(M_{ZP}+1) \rightarrow M_T$, wrong synchronization is performed and the false alarm probability (P_F) increases:

IV. EVALUATION

A. Simulation

In simulation of the proposed algorithm (ESTS), it is assumed that there are no other imperfections except timing offset. 100 realization of channel model CM1 (0-4 meter line of sight and 5 nanosecond delay spread) and CM2 (0-4 meter non line of sight and 8 nanosecond delay spread) are considered in simulation. It is also assumed that the first pattern of time-frequency code (TFC1) is used in data transmission and frequency synchronization is ideal. The performance of the system is evaluated by the probability of synchronization (P_{sync}), bit error rate (BER) and the MSE of timing offset as defined below.

$$MSE = \sum_{\forall \theta} \left(\theta - \hat{\theta}\right) P_{sync}\left(\hat{\theta}\right) \quad (16)$$

Where $P_{sync}\left(\hat{\theta}\right)$ is the probability of synchronization at $\hat{\theta}$ for the simulated channel realization and $P_F + P_{sync} = 1$.

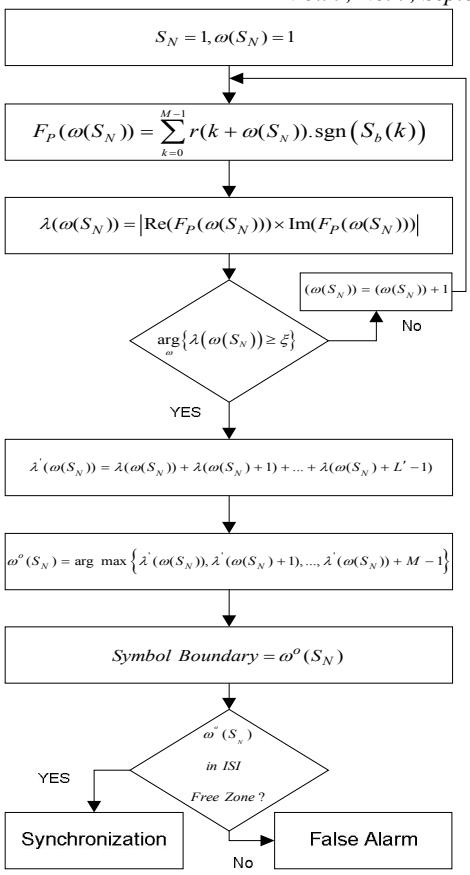


Figure 3. Flowchart of Proposed ESTS algorithm

The threshold value which is used in coarse synchronization is defined so we have low MSE and high P_{sync} . By simulation results, threshold value is considered to be 24 dB and 23 dB for CM1 and CM2 respectively.

We also need to define the number of required cross correlations to minimize the effect of delay spread in multipath fading channels. For a given threshold at a certain SNR, the MSE decreases while the P_{sync} increases when L' increases up to 15 and the performance measures stay constant afterwards. So we consider the $L' = 15$ as the number of required cross correlations. Simulation results for the MSE and P_{sync} metrics are shown in "Fig 4" and "Fig 5" respectively. As shown in "Fig 4" in the MSE metric, a great improvement is achieved in all SNR values especially in low values both in CM1 and CM2 channel model compared with those of the CBTS and FTA. "Fig 5" indicates that in P_{sync} metric and high SNR values, the performance is the same as that of the CBTS algorithm in CM1 channel. In low SNR values and both CM1 and CM2 channel models and high SNR values in CM1 channel model, performance is improved compared with that of the CBTS. In all SNR values and both channel models, performance of the proposed algorithm is better than that of the FTA.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In "Fig 6" and "Fig 7" the bit error rate of the proposed algorithm is compared with those of, [6-7] in CM1 and CM2 channel model, respectively.

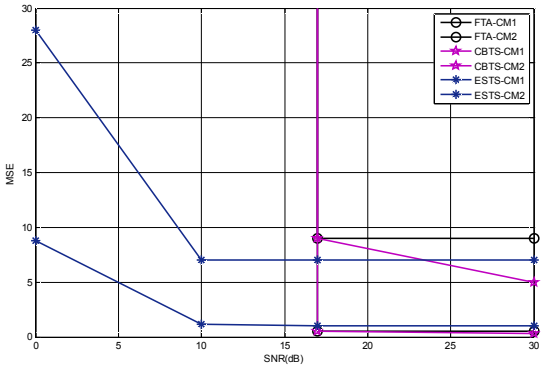


Figure 4. Comparison of MSE for proposed algorithm (ESTS), FTA [6] and CBTS [7] in CM1 and CM2 channel models.

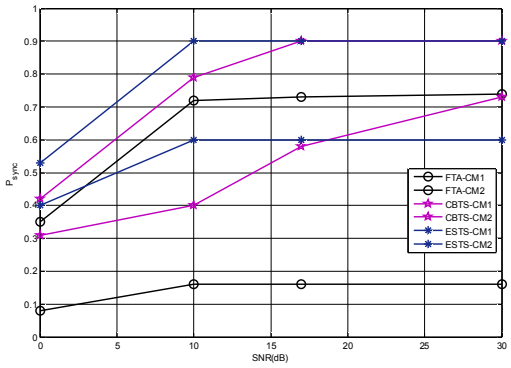


Figure 5. Comparison of Psync for proposed algorithm (ESTS), FTA [6] and CBTS [7] in CM1 and CM2 channel models.

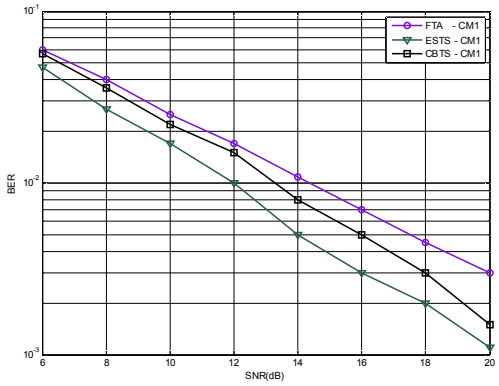


Figure 6. Comparison of BER for proposed algorithm (ESTS), FTA [6] and CBTS [7] in CM1 channel model.

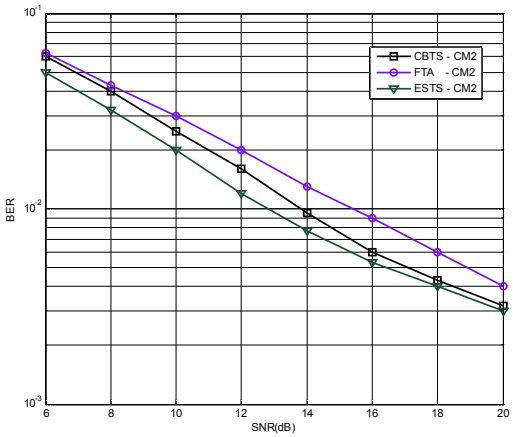


Figure 7. Comparison of BER for proposed algorithm (ESTS), FTA [6] and CBTS [7] in CM2 channel model.

B. *Computational Complexity*

To compare the computational complexity, we assume that there are no only pure noise packets as considered in [6] and [7]. So, we skip the coarse synchronization part (packet detection). We also assume that the recursive "(14)" is used instead of "(13)". According to [6] and [7] the number of multiplications in FTA, CBTS and proposed algorithm are $(5M_T+\theta)\times(2M+1)$, $(5M_T+\theta)\times(2M-1)$ and $M\times(L'+M-1)$ respectively. The numbers of summations are also $(5M_T+\theta)\times(2M-2)$, $(5M_T+\theta)\times(2M)$ and $((M-1)\times(M+L'+1)+L'-1)$ in the same order. As a numerical result, by considering $M=128,\ M_{ZP}=32$ and $M_g=5,\ M_T=165,\ \theta=1\ and\ L'=15$, the number of multiplications in the FTA, CBTS and proposed algorithm (ESTS) are 212282, 210630 and 18176, respectively, which show that the proposed algorithm is less complex. In the same order, the numbers of summations are equal to 209804, 211456 and 18302.

V. Conclusion

In this paper, a new efficient symbol timing synchronization (ESTS) algorithm proposed for MB-OFDM UWB systems. In the proposed algorithm, was compared in MSE, synchronization probability and bit error rate metrics with those of [6] and [7]. Simulation results show a great improvement while the computational complexity is reduced.

References

[1] ECMA-368, "High Rate Ultra Wideband PHY and MAC standard, 3rd Edition" December 2008, http://ecmainternationa.org/publications/files/ECMA-ST/ECMA-368.pdf.

[2] J. Van De Beek, M. Sandell and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," IEEE Trans. Signal Processing, vol.45, no. 7, pp.1800-1805, July 1997.

[3] Guo Yi, Liu Gang and Ge Jianhua, "A novel time and frequency synchronization scheme for OFDM systems," IEEE Transactions on Consumer Electronics. vol.54, no.2, pp.321-325, 2008.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[4] Juan I. Montojo and Laurence B. Milstein, "Effect of Imperfections on the performance of OFDM Systems" IEEE Trans. COMMUNICATIONS, vol. 57, no. 7, 2009.

[5] C. W. Yak, Z. Lei, S. Chattong and T. T. Tjhung, "Timing synchronization for Ultra-Wideband (UWB) Multi-Band OFDM systems," IEEE Vehicular Technology Conference. Germany, pp.1599-1603, 2005.

[6] Debarati Sen, Saswat Chakrabarti, and R. V. Rajakumar, "Symbol timing synchronization for Ultra-Wideband (UWB) Multi-band OFDM (MB-OFDM) systems," IEEE COMSWARE. India.pp. 5-10, 2008.

[7] Xiaoyan Wang, Zaichen Zhang and Chuanyou Chen," A Robust Time Synchronization Scheme for MB-OFDM UWB System," IEEE International Conference on Signal Processing Systems. Dalian, pp.529-532, 2010.

[8] M.N. Suresh, D.C. Mouli, M.G Prasath, V. Abhaikumar and S.J. Thiruvengadam, "Symbol timing estimation in multiband OFDM based ultra wideband system," IEEE International Conference on Signal Processing and Communications, India, pp. 1–4, 2010.

[9] Sen, S. Chakrabarti, and R.V. Raja Kumar, "A new timing estimation and compensation scheme for ultra-wideband MB-OFDM communications," in Proceedings of IEEE WOCN, May 2008, pp. 1-5.

[10] J. Foerster, "Channel modeling sub-committee report final," IEEE, Document IEEE P802.15-02/490r1-SG3a, 2003.

[11] Wimedia alliance, multiband OFDM physical layer specification, release 1.1 July 14, 2005.

[12] A.F. molish, "Ultra wideband propagation channel theory, measurement, and modeling," IEEE Transactions on Vehicular Technology, Sept. 2005, vol.54, no.5, pp.1528-1545.

[13] S. Johansson, M. Nilsson, and P. Nilsson, "An OFDM timing synchronization ASIC," in Proceedings of IEEE ICECS, December 2000, pp. I.324-I.327.

[14] Dardari and M. Z. Win, "Threshold-based time-of-arrival estimators in UWB dense multipath channels," in Proceedings of IEEE ICC, June 2006, pp. 4723-4728.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Securing the Multilevel Information System

MOHAN H.S.
Research Scholar
Dr. MGR University, Chennai, INDIA

A.RAJI REDDY
Professor & Head, Dept of ECE,
Madanapalle Institute of Technology & Science,
Madanapalle, Chittoor, INDIA
.

Abstract— **Nowadays, multilevel secure database is common in distributed systems. These databases require a generalized software system for multiuser and simultaneous access in the distributed system, as the client systems may be dissimilar (heterogeneous hardware and software.) The information system will usually be a blend of both information retrieval system and information management (create and maintain) system. This paper gives an approach in developing a generalized multilevel secure information system using three-tier architecture. The approach shows how data level integrity can be achieved using access and security levels on users/subjects and data/objects respectively.**

Keywords- multilevel secure database; information system; generalized software system

I. INTRODUCTION

The continuing growth of essential data is leading to the popularity of databases and database management system. A database is a collection of related data. Database management system (DBMS) is a collection of programs that enable users to create and maintain a database. A good database management system generally has the ability to protect data and system resources from security breaches like intrusions, unauthorized modification, unauthorized copying and observation, etc [2]. Damage to the important data will not only affect a single user or application, but the entire information system and the corporation will be affected. Secrecy and integrity of data are of major concern in information system while handling the data. Secrecy means preventing unauthorized users from copying and observation while retrieving data. Integrity means preventing unauthorized users from creating, modifying and deleting the data.

In a multilevel secure database, the data is assigned with security levels for attaining secrecy and integrity [2]. Everyone cannot access all the data in such a database. This database exists in a distributed system and is simultaneously accessed by multiple users. This requires a generalization of software system that enables multiple users to simultaneously access the multilevel secure database.

The new approach uses the three-tier architecture [4] to develop a software system that allows users of different levels to retrieve, create and maintain data simultaneously. The authentication of users is handled both at client end as well as the server end, which ensures high security. The approach uses multilevel secure data model at the database and multilevel users to access the data. The classification of data/objects and users/subjects has been done in two ways –top secure model and secure model. The users have been categorized into View only (V) users and Privileged (P) users. The view only user's access levels have been categorized into Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U.) The privileged user's access levels have been categorized into two hierarchical levels –the first being Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U) and the second level being create-modify (CM) and create-modify-delete (CMD). The top secure model uses the both the hierarchical levels of classification for privileged user. The secure model uses only first level of hierarchical classification for privileged user. The access levels for view only user is same for both –top secure model and secure model. The configurable data elements are classified into Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U.) The classification of data/object is given in detail in section 3. With the levels defined for both, users and data, the approach proceeds in achieving such a software system. This approach helps in the development of a multilevel secure information system.

The remaining part of the paper is organized as follows. Section 2 gives a brief description of related work carried out in this direction. Section 3 describes the new approach. Section 4 gives the implementation of this approach in a simple distributed system using Java. Section 5 discusses the advantages of the said approach. Section 6 discusses the limitations of said approach and section 7 concludes.

II. RELATED WORK

Different authors have given different types of multilevel relational data model until now. Some of the related scenarios are as discussed next. Sea View is a multilevel relational data model, developed in the context of the Sea View project [3, 6]. The Sea View project is a joint project by SRI International and Gemini Computers, Inc. The project also defined MSQL, an extension of SQL to handle multilevel data. The Sea View security model consists of two components –the MAC (Mandatory Access Control) model and TLB (Trusted Computing Base) model [6]. The MAC model defines the mandatory security policy. Each subject is assigned a readclass and a writeclass. A subject can read an object if the subject's readclass dominates the access class of the object. A subject can write into an object if the object's class dominates the writeclass of the subject. The TCB model defines discretionary security and supporting policies for multilevel relations, views,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and integrity constraints, among others. The data model on which Sea View is based is a multilevel relational data model. Multilevel relations are implemented as views over single level relations, that is, over relations having a single access class associated with them.

Jajodia and Sandhu proposed a reference model for multilevel relational DBMSs and addressed on a formal basis entity integrity and update operations in the context of multilevel databases [7]. In the model by Jajodia and Sandhu a multilevel relation schema is denoted as R(A1,C1,…….,An,Cn,TC), where Ai is an attribute over a domain Di, and Ci is a classification attribute for Ai, i = 1,…,n. The domain of Ci is the set of access classes that can be associated with attribute Ai. TC is the classification attribute of the tuples. Furthermore, for each access class c, a relation instance Rc is defined. Elements of Rc are of the form R(a1,c1,….an,cn,tc), where ai is a value in the domain Di, ci is a classification attribute for ai, i =1,…..,n, and tc is the classification attribute of the tuples; tc is determined by computing the least upper bound of each ci in the tuple. The relation instance Rc represents a view of the multilevel relation for subjects having access class c. The instance at level c is obtained from the multilevel relation by masking all attribute values whose classification s higher than or incomparable with c. This is obtained by substituting them with null values. Thus, subjects with different access classes have different views of the same multilevel relation data model is restated as follows: a multilevel relation R satisfies the entity integrity property if, for all instances Rc of R, and for each tuple t of Rc, the following conditions are satisfied:

a) The attributes of the primary key must be not null in t;

b) The attributes of the primary key must have the same access class in t;

c) The access class associated with a nonkey attribute must dominate the access classes associated with the attributes in the primary key.

The model by Jajodia and Sandhu supports both attribute and tuple polyinstantiation. Similar to the Sea View model [3, 6], the key of a multilevel relation is defined as a combination of attributes, their classifications, and the classification of all the other attributes in the relation.

The Multilevel Relational (MLR) data model proposed by Chen and Sandhu in [8] is an extension of the model proposed by Jajodia and Sandhu [7]. The data model is basically the one presented in previous paragraph, the main difference being that in the MLR data model the constraint that there can be at most one tuple in each access for a given entity is imposed. The MLR model tries to overcome some of the ambiguities contained in the Jajodia and Sandhu model. In the MLR model a new semantics for data classified at different levels is proposed, based on the following principles:

a) The data accepted by a subject at a given security level consist of two parts: (i) the data classified at his/her level and (ii) the data borrowed from lower levels;

b) The data a subject can view are those accepted by subjects at his/her level and by subjects at lower levels;

c) A tuple with classification attribute c contains all the data accepted by subjects of level c.

III. MULTI-LEVEL SECURITY

A generalization of software system (for information system) that enables multiple users to simultaneously access, create and maintain (insert, update, delete) can be achieved by using a three-tier architecture. A software system in a distributed system using three-tier architecture must have three components –clients, server and database. The database system used may be an open source or commercial systems. In three-tier architecture the client systems can be dissimilar but the generalization of software systems achieves single application specific server for all these clients.

Fig. 1 shows the three-tier architecture. The database will be a shared resource among all clients using the software system. The client software can be written using any programming language but the clients must have the knowledge of communicating with the server. The application specific business rules (procedures, constraints) are stored at server. The server ensures the identity of the client and accesses the data from the database on behalf of client [5]. In this way even in a distributed system the business rules can be common for all clients requesting the data from server. The generalization can be achieved by the development of the middle-tier i.e., server. Any upgradation in a business rule or a database change requires upgradation only in server and do not affect the client softwares in that system.

Fig. 2 describes how the security levels can be expressed as a linear order with four security levels: Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U.) Partial ordering has been omitted intentionally to make the model less complicated.

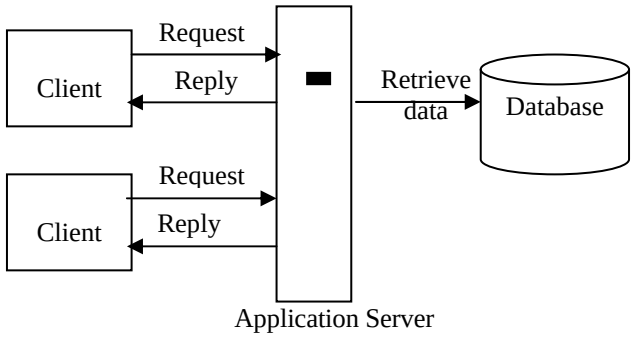


Figure 1. Three-tier Architecture

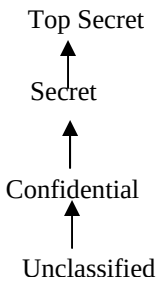


Figure 2. Security levels in linear order

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. APPROACH FOR MULTI LEVEL SECURE INFORMATION SYSTEM

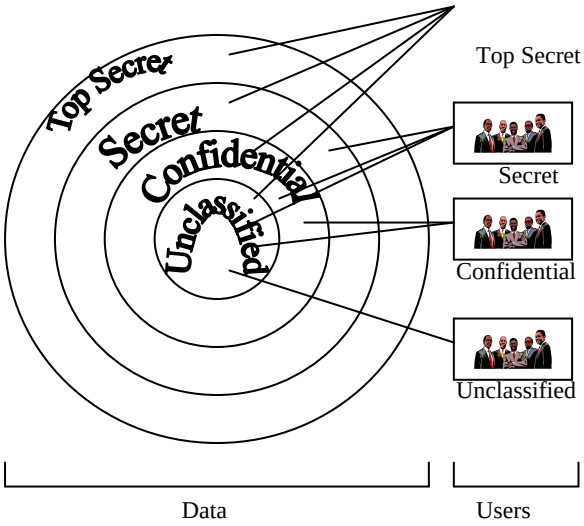


Figure 3. Various users accessing data at various security levels

Fig. 1, three-tier architecture, when observed indicates that to make such an information system multilevel secured, the clients and data in database, both must be classified at various levels. These levels together define the levels for security in an information system.

First let us classify the clients. The users have to be categorized into View only (V) users and Privileged (P) users. The view only user can just retrieve the data but he cannot modify the data. The privileged user can both retrieve and maintain the data. The view only user's access levels have been categorized into Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U.) The privileged user's access levels have been categorized into two hierarchical levels –the first being Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U) and the second level being create-modify (CM) and create-modify-delete (CMD). Finally the classification or access levels of users can be in two forms: {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,CM,) (P,S,CM,) (P,C,CM,) (P,U,CM,) (P,TS,CMD,) (P,S,CMD,) (P,C,CMD,) (P,U,CMD)} and {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,) (P,S,) (P,C,) (P,U)}.

Secondly, the data in database must be classified. The configurable data elements are classified into Top Secret (TS,) Secret (S,) Confidential (C) and Unclassified (U.) The multilevel relation schema 'R' can be denoted in two forms as R(A1,C1,A2,C2, A3,C3,…….,An,Cn,TC) and R(A1,A2,A3,…….,An,TC), where Ai is an attribute over a domain Di, and Ci is a classification attribute for Ai, i = 1,…,n. The domain of Ci is the set of access classes {Top secret (TS,) Secret (S,) Confidential (C,) Unclassified (U)} or {Ci} that can be associated with attribute Ai and defines security level of the attribute. TC (tuple classification) is the classification attribute of the tuples and takes value {TS,S,C,U} to define the security level of the tuple.

The combination of the above two classifications (users and data) give rise to four various ways in which an information system can be made multilevel secured. The two models, used to achieve secrecy and integrity of data in information system are –top secure model and secure model. They are as discussed below.

A. Top Secure model

Case 1: High multilevel security (some attributes must be accessible by certain level users) is needed for data and high multilevel access for users.

The two components to be implemented are multilevel relational data model and access control. Multilevel relational data model used for top secure model is as follows: the multilevel relation schema is denoted as R(A1,C1,…….,An,Cn,TC), where Ai is an attribute over a domain Di, and Ci is a classification attribute for Ai, i = 1,…,n. The domain of Ci is the set of access classes {Top secret (TS,) Secret (S,) Confidential(C,) Unclassified (U)} or {Ci} that can be associated with attribute Ai. TC is the classification attribute of the tuples and takes value {TS,S,C,U.}

The users are classified as {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,CM,) (P,S,CM,) (P,C,CM,) (P,U,CM) (P,TS,CMD,) (P,S,CMD,) (P,C,CMD,) (P,U,CMD)} described above. If user/password authentication scheme [5] is used to achieve this user classification then the schema for the multilevel relation user can be R(userid, username, password, viewLevel, accessLevel, updateLevel) where viewLevel takes the value {V,P,} accessLevel takes the value {TS,S,C,U,} and updateLevel takes the value {C,CMD.} Fig. 4 and Fig. 5 show the top-secret and secret instances for an example of top secure model.

Employee	C1	Job	C2	Salary	C3	TC
Laxmi	S	Architect	S	20K	TS	TS
Vidya	TS	Agent	TS	17.5K	TS	TS
Parvathi	U	AT	U	8K	C	C
Priya	U	PT	U	Null	U	U
Lolitha	C	Sr. Engi.	S	19K	S	S

Figure 4. Top-Secret Instance for Top Secure model

Employee	C1	Job	C2	Salary	C3	TC
Laxmi	S	Architect	S	Null	TS	TS
Parvathi	U	AT	U	8K	C	C
Priya	U	PT	U	Null	U	U
Lolitha	C	Sr. Engi.	S	19K	S	S

Figure 5. Secret Instance for Top Secure model

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Case 2: High multilevel security (some attributes must be accessible by certain level users) is needed for data and multilevel access for users.

Multilevel relational data model used for top secure model is as follows: the multilevel relation schema is denoted as R(A1,C1,……,An,Cn,TC), where Ai is an attribute over a domain Di, and Ci is a classification attribute for Ai, i = 1,…,n. The domain of Ci is the set of access classes {Top secret (TS,) Secret (S,) Confidential(C,) Unclassified (U)} or {Ci} that can be associated with attribute Ai. TC is the classification attribute of the tuples and takes value {TS,S,C,U.}

The users are classified as {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,) (P,S,) (P,C,) (P,U,)} described above. If user/password authentication scheme [5] is used to achieve this user classification then the schema for the multilevel relation user can be R(userid, username, password, viewLevel, accessLevel) where viewLevel takes the value {V,P,} and accessLevel takes the value {TS,S,C,U.}

B. Secure Model

Case 1: Multilevel security (some attributes must be accessible by certain level users) is needed for data and high multilevel access for users.

The two components to be implemented are multilevel relational data model and access control. Multilevel relational data model used for secure model is as follows: the multilevel relation schema is denoted as R(A1,……,An,TC), where Ai is an attribute over a domain Di, i = 1,…,n. TC is the classification attribute of the tuples and takes value {TS,S,C,U.}

The users are classified as {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,CM,) (P,S,CM,) (P,C,CM,) (P,U,CM) (P,TS,CMD,) (P,S,CMD,) (P,C,CMD,) (P,U,CMD)} described above. If user/password authentication scheme [5] is used to achieve this user classification then the schema for the multilevel relation user can be R(userid, username, password, viewLevel, accessLevel, updateLevel) where viewLevel takes the value {V,P,} accessLevel takes the value {TS,S,C,U,} and updateLevel takes the value {C,CMD.} Fig. 6 and Fig. 7 show the top-secret and secret instances for an example of secure model.

Employee	Job	Salary	TC
Laxmi	Architect	20K	TS
Vidya	Agent	17.5K	TS
Parvathi	AT	8K	C
Priya	PT	Null	U
Lolitha	Sr. Engi.	19K	S

Figure 6. Top-Secret Instance for Secure model

Employee	Job	Salary	TC
Parvathi	AT	8K	C
Priya	PT	Null	U
Lolitha	Sr. Engi.	19K	S

Figure 7. Secret Instance for Secure model

Case 2: Multilevel security (some attributes must be accessible by certain level users) is needed for data and multilevel access for users. A successful implementation of multilevel secured information system of this category has been described [1] in the paper.

Multilevel relational data model used for secure model is as follows: the multilevel relation schema is denoted as R(A1,……,An,TC), where Ai is an attribute over a domain Di, i = 1,…,n. TC is the classification attribute of the tuples and takes value {TS,S,C,U.}

The users are classified as {(V,TS,) (V,S,) (V,C,) (V,U,) (P,TS,) (P,S,) (P,C,) (P,U,)} given in section 1. If user/password authentication scheme is used to achieve this user classification then the schema for the multilevel relation user can be R(userid, username, password, view level, access Level) where view level takes the value {V,P,} and access Level takes the value {TS,S,C,U.}

The top secure model uses both the hierarchical levels of classification for privileged user. The secure model uses only first level of hierarchical classification for privileged user. The access levels for view only user is same for both –top secure model and secure model. The point to be observed in both the models –top secure and secure is that instantiation is omitted. There will be only one tuple (considering TC) whose security level will be {TS, S, C, U} or the tuple will not exist. If it already exists at higher security level like TS, then it is not viewable by users at lower access levels {S, C, U} and they are not be permitted to even create another tuple with same primary key.

Fig. 3 describes how the users are related to data. Now let us define the rules for using the top secure and secure model. The rules can be given as follows:

Rule 1: The attributes of the primary key must be not null.

Rule 2: The attributes of the primary key must have the same security level in a tuple t.

Rule 3: The security level of the attributes of primary key must be either at the same level as TC or at lower levels in a tuple t.

Rule 4: The security level associated with a nonkey attribute must be either at the same level as TC or at lower levels in a tuple t.

Rule 5: The data accepted by a user at a given security level consist of two parts: (i) the data classified at his/her level; and (ii) the data borrowed from lower levels.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Rule 6: The data a user can view are those accepted by users at his/her level and by subjects at lower levels.

Rule 7: A tuple with classification attribute (TC) c contains all the data accepted by users of level c (includes lower levels,) where c ={TS, S, C, U.}

Rule 8: The configurable data elements take only one value at any time and their value is accessible by users at the same level or higher. But its value to users at lower levels is null (ambiguity, i.e. does not exist or not available) or not accessible depending on the developed information system.

The classified data must not only be protected from direct access by unauthorized users, but also from disclosure through indirect means, such as inference. For example, a low user attempting to access a high object can infer something depending upon whether the system responds with "object not found" or "permission denied."

With the top secure model and secure model defined we now proceed towards using them in the information system to achieve multilevel security. The common components in these two models are the implementation of multilevel secure data model and access control. The selection of the type is left to programmer's choice based on requirements. In the three-tier architecture used certain rules have to be followed: the multilevel security data model has to be implemented in the database and access control has to be implemented on all the three components, that is, client graphical user interfaces (GUIs), server and database. The requirements apart from rules are as follows.

Req 1: The server has to keep a check on all requests for view level and access level in case of secure model and also update level in case of top secure models.

Req 2: The server has to maintain session details and control the users who have currently logged to ensure security [5].

Req 3: The communication between clients and server has to be secured from intrusions according to the requirement.

Req 4: One of the top secure or secure model has to be implemented on the data in database and users, in the distributed system according to the requirement.

Fig. 8 shows the model for access control using three-tier architecture. The reference monitor grants or denies access for various access requests from different users. The very nature of 'access' suggests that there is an active subject accessing a passive object with some specific access operation, while a reference monitor grants or denies access. The reference monitor is present within the application server responsible for controlling access. This approach can be extended and implemented for N-tier architecture where N is more than 3. But the data manager or application server handling data at the database is common for all N-tier architecture. Thus, a single reference monitor handles all the access requests. As each access is secured the whole system is said to be secure (basic security theorem.) Securing the data is not only protecting the data from direct access by unauthorized users, but also from disclosure through indirect means, such as covert signaling channels and inference.

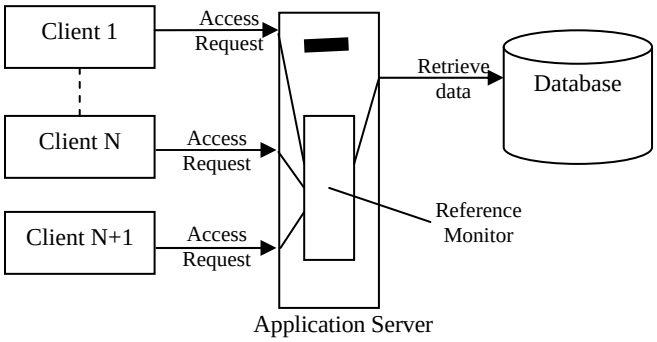


Figure 8. Model of access control

The application server with the reference monitor must ensure that systems connecting it are trusted computers. To achieve generalization of information system using three-tier architecture [1] the following procedure is to be followed. The database must be a shared resource among all clients using the information system. The client systems are dissimilar but there are no rules to be followed unlike two-tier architecture. The client software can be written using any programming language but the clients must have the capability of communicating with the server (reference monitor.) The application specific business rules (procedures, constraints) are stored at application server (consisting of reference monitor.) The application server ensures the identity of the client and accesses the data from the database on behalf of client. In this way even in a distributed system the access rules can be common for all clients requesting the data from application server. The generalization can be achieved by the development of the middle-tier i.e., application server. Any upgradation in an access rule or a database change requires upgradation only in application server and do not affect the client softwares using that software system. Hence, three-tier architecture is more suitable for the generalization of information system.

The design of middle tier i.e., application server requires monitoring through various issues [5]: connectionless vs. connection-oriented server access, stateless vs. stateful applications, and iterative vs. concurrent server implementations. The most suitable approach can be taken based on the type of environment and application.

With multiple clients accessing and potentially modifying the shared data or information, maintaining the integrity of the data or information will be an important issue. The application server must consist of a mediator who monitors the shared data or information for maintaining its integrity. The mediator can use locking techniques for the same. Once a change occurs the updates can be broadcast. It will not become a bottleneck when the size of the system scales up because we are discussing it with respect to multilevel secure information system. The information system that requires multilevel security will not have very huge size so as to create a bottleneck. Moreover the approach recommends a separate application server for each database. With this each database will have a separate application server to handle the access requests in a distributed

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

environment. Thus, we use three-tier architecture to render the system generalized.

If an information system has been implemented using the above given approach (with all the requirements from 1-4 and rules 1-8 implemented,) then the information system in a distributed environment is considered to be generalized and multilevel secured.

V. ADVANTAGES OF GIVEN APPROACH

The specified approach has many advantages of using it and they are given below.

A. Security

The specified approach ensures data integrity. It uses top secure model or secure model (according to information system requirement) for implementing multilevel security. The multilevel for users and data go a long way in securing the information system.

B. Encapsulation of services and data

The given approach uses three-tier architecture where all the services reside on server and the server also masks the location of data. Encapsulation of data is achieved, as the clients do not know the schema structure of the stored data.

C. Administration

In three-tier architecture used all the client applications accessing data are centrally managed on the server, which results in cheaper maintenance and less complex administration of information system [4].

D. Flexibility in the approach

The given approach is just a generic approach and can be used in the implementation of various information systems. The communication between the client and server should be secure, but what type of security is to be provided is decided by the developer based on his requirements. Hence the approach is a common approach for many systems, but when the rules 1-8 (+ Req 1-4 fulfilled) are followed any information system becomes multilevel secured.

E. Application reuse

The specified approach encapsulates the data and services at the server. The server can reuse services and objects but an added advantage is the legacy application integration is possible through gateways encapsulated by services and objects.

F. Generalization

The given approach generalizes the software for the information system, as the business rules are stored at server and the clients using services from server can be using different platforms, hardware and softwares

VI. LIMITATIONS OF GIVEN APPROACH

A. Ease of development

The specified approach requires hybrid skills that include transaction processing, database design, communication experience, graphical user interface design, etc. The more advanced applications require knowledge of distributed objects and component infrastructures [4]. But the ease of development is only getting better with standard tools emerging.

B. Instantiation unused

The given approach avoids the use of instantiation in its approach that deprives the approach of the advantages of instantiation. But the disadvantages of using instantiation are also avoided.

VII. CONCLUSION

This paper gives a novel approach towards making the information system multilevel secured and generalized. This paper gives an explanation of the same and discusses its advantages and drawbacks.

ACKNOWLEDGMENT

The author would like to thank Dr. A Raji Reddy for his continuous support and guidance for carrying the research work.

REFERENCES

[1] C.N. Deepika and W.V. Eswaraprakash, *"Interoperable Three-Tier Database Model,"* The Journal of Spacecraft Technology, Vol. 17, No. 2, July 2007, pp.16-22.

[2] Ramez Elmasri and Shamkant B. Navathe, *Fundamentals Of Database Systems*, 3rd ed., Pearson Education, Asia, 2002, pp.715-726.

[3] Teresa F. Lunt, *Research Directions in Database Security*, Springer-Verlag, New York, 1992, pp.13-31.

[4] Robert Orfali, Dan Harkey, Jeri Edwards, *The Essential Client/Server Survival Guide*, 2nd ed., John Wiley & Sons, U.S.A., 1996, pp.19-20.

[5] Douglas E. Comer and David L. Stevens, *Internetworking With TCP/IP*, vol. 3.: *Client-Server Programming And Applications*, Prentice-Hall, U.S.A., 1993.

[6] D. E. Denning, T. F. Lunt, R. R. Schell, M. Heckman and W. Shockley, *"A Multilevel Relational Data Model,"* In Proc. of the IEEE Symposium on Security and Privacy, Oakland, C. A., April 1987, pp.220-234.

[7] Jajodia S. and Sandhu R. S., "Toward a Multilevel Secure Relational Data model," In Proc. of ACM Sigmod International Conference on Management of Data, Denver, C. O., May 1991, pp.50-59

[8] Chen F. and Sandhu R. S., *"The semantics and expressive power of the MLR data model,"* In Proc. of the IEEE Symposium on Security and Privacy, Oakland, C. A., May 1995, pp.128-142.

AUTHORS PROFILE

Mohan H.S. received his Bachelor's degree in computer Science and Engineering from Malnad college of Engineering, Hassan during the year 1999 and M. Tech in computer Science and Engineering from Jawaharlal Nehru National College of Engineering, Shimoga during the year 2004. Currently pursing his part time Ph.D degree in Dr. MGR university, Chennai. He is working as

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

a professor in the Dept of Information Science and Engineering at SJB Institute of Technology, Bangalore-60. He is having total 13 years of teaching experience. His area of interests are Networks Security, Image processing, Data Structures, Computer Graphics, finite automata and formal languages, Compiler Design. He has obtained a best teacher award for his teaching during the year 2008 at SJBIT Bangalore-60. He has published and presented papers in journals, international and national level conferences

A. Raji reddy received his M.Sc from Osmania University and M.Tech in Electrical and Electronics and communication Engineering from IIT, Kharagpur during the year 1979 and his Ph.D degree from IIT, kharagpur during the year 1986.He worked as a senior scientist in R&D of ITI Ltd, Bangalore for about 24 years. He is currently working as a professor and head in the department of Electronics and Communication, Madanapalle Institute of Technology & Science. Madanapalle. His current research areas in Cryptography and its application to wireless systems and network security. He has published and presented papers in journals, international and national level conferences.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Streamed Coefficients Approach for Quantization Table Estimation in JPEG Images

Salma Hamdy
Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
s.hamdy@cis.asu.edu.eg

***Abstract*— A forensic analyst is often confronted with low quality digital images, in terms of resolution and/or compression, raising the need for forensic tools specifically applicable to detecting tampering in low quality images. In this paper we propose a method for quantization table estimation for JPEG compressed images, based on streamed DCT coefficients. Reconstructed dequantized DCT coefficients are used with their corresponding compressed values to estimate quantization steps. Rounding errors and truncations errors are excluded to eliminate the need for statistical modeling and minimize estimation errors, respectively. Furthermore, the estimated values are then used with distortion measures in verifying the authenticity of test images and exposing forged parts if any. The method shows high average estimation accuracy of around 93.64% against MLE and power spectrum methods. Detection performance resulted in an average false negative rate of 6.64% and 1.69% for two distortion measures, respectively.**

Keywords: Digital image forensics; forgery detection; compression history; Quantization tables.

I. INTRODUCTION

Most digital image forgery detection techniques require the doubtful image to be uncompressed and in high quality. Yet, currently most acquisition and manipulation tools use the JPEG standard for image compression. JPEG images are the most widely used image format, particularly in digital cameras, due to its efficiency of compression and may require special treatment in image forensics applications because of the effect of quantization and data loss. Usually JPEG compression introduces blocking artifacts and hence one of the standard approaches is to use inconsistencies in these blocking fingerprints as a reliable indicator of possible tampering [1]. These can also be used to determine what method of forgery was used. Moreover, a digital manipulation process usually ends in saving the forgery also in JPEG format creating a double compressed image. Mainly, two kinds of problems are addressed in JPEG forensics; detecting double JPEG compression, and estimating the quantization parameters for JPEG compressed images. Double compressed images contain specific artifacts that can be employed to distinguish them from single compressed images [2-4]. Note, however, that detecting double JPEG compression does not necessarily prove malicious tampering: it is possible, for example, that a user may re-save high quality JPEG images with lower quality to save storage space. The authenticity of a double JPEG compressed image, however, is at least questionable and further analysis would be required. Generally, the JPEG artifacts can also be used to determine what method of forgery was used. Many passive schemes have been developed based on these fingerprints to detect re-sampling [5] and copy-paste [6-7]. Other methods try to identify bitmap compression history using Maximum Likelihood Estimation (MLE) [8-9], or by modeling the distribution of quantized DCT coefficients, like the use of Benford's law [10], or modeling acquisition devices [11]. Image acquisition devices (cameras, scanners, medical imaging devices) are configured differently in order to balance compression and quality. As described in [12-13], these differences can be used to identify the source camera model of an image. Moreover, Farid [14] describes JPEG *ghosts* as an approach to detect parts of an image that were compressed at lower qualities than the rest of the image and uses to detect composites. In [15], we proposed a method based on the maximum peak of the histogram of DCT coefficients.

Furthermore, due to the nature of digital media and the advanced digital image processing techniques, digital images may be altered and redistributed very easily forming a rising threat in the public domain. Hence, ensuring that media content is credible and has not been altered is becoming an important issue governmental security and commercial applications. As a result, research is being conducted for developing authentication methods and tamper detection techniques.

In this paper, we propose an approach for quantization table estimation for single compressed JPEG images based on streamed DCT coefficients. We show the efficiency of this approach and how it recovers the weak performance of the method in [15] for high quality factors.

In section 2 we describe the approach used for estimating quantization steps of JPEG images, and the two distortion measures we use in our forgery detection process. Experimental results are discussed in section 3. Section 4 is for conclusions. A general model for forgery detection based on quantization table estimation is depicted in Fig. 1.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

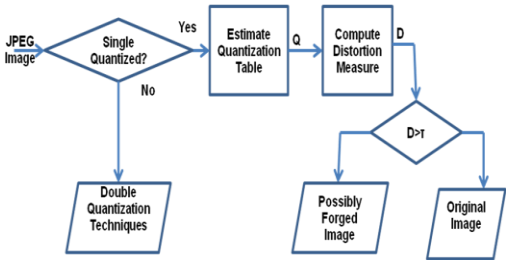


Figure 1. A general model for forgery detection using quantization tables.

II. STREAMED COEFFICIENTS APPROACH

In [15] we proposed an approach for estimating quantization tables for single compressed JPEG images based on the absolute histogram of reconstructed DCT coefficients. Since we could not use the "temporary" values of the dequantized coefficients X_q to build the histograms, We managed to reverse the process one step, i.e. to undo the IDCT, and reconstruct the coefficients by taking the block DCT of the decompressed image and compensate for errors (Fig. 2). This "re-compression" step produces an estimate X^* that we used in our maximum peak method in [15].

Now, if we continue one step further in reverse, that is, undo the dequantization, the normal case requires the quantization table to compress and reach the final version of the coefficients that are encoded and dumped to the file. However, the quantization table is unknown and it is our goal to estimate it. Yet, we have the result of the quantization; the compressed coefficients, which we can retrieve from the file, as shown in Fig. 3. Hence, we can conclude a straightforward relation between the streamed compressed coefficients, and the reconstructed dequantized DCT coefficient. If we refer to the decompressed image as I, then we have:

$$I = IDCT(X_q) = IDCT[DQ(X_s)] \qquad (1)$$

where DQ is the dequantization process, and X_s resembles the compressed coefficient dumped from the image file. As we pointed out above, the dequantized coefficient can be estimated (reconstructed) through applying the inverse of this step which is the discrete cosine transform. Hence:

$$\begin{aligned} DCT(I) &= DCT[IDCT(X_q)] \\ &= DCT[IDCT[DQ(X_s)]] \\ X_q &= DQ(X_s) \end{aligned} \qquad (2)$$

Again, X_q is only temporary and is evaluated as its reconstructed copy X^* taking into consideration the error caused by the cosine transforms. Hence, (2) becomes:

$$X^* \pm E = DQ(X_s) \qquad (3)$$

where E is the error caused by the cosine transforms. Since a compressed coefficient is dequantized via multiplying it by the corresponding quantization step we can write:

$$X^* \pm E = qX_s \qquad (4)$$

Finally, solving for q gives:

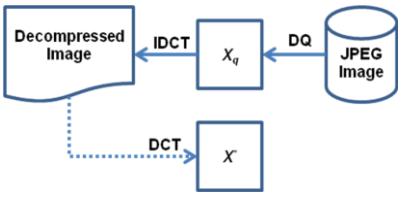


Figure 2. X_q is an intermediate result. Taking the DCT of a decompressed image block does not reproduce X_q exactly, but an approximation to it; X^*.

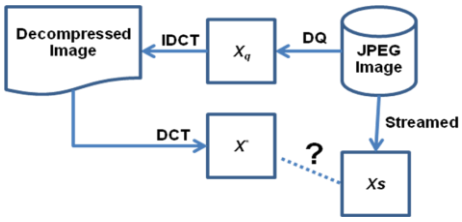


Figure 3. X_q is an intermediate result. Taking the DCT of a decompressed image block does not reproduce X_q exactly, but an approximation to it; X^*.

$$q = \frac{X^* \pm E}{X_s} \qquad (5)$$

Again we suggest the neglect of round off errors; as we see their effect could be minimal and could be compensated for using lookup tables if needed, also the exclusion of saturated blocks to minimize the possibility of truncation errors. Hence, the estimated quantization step is computed as:

$$q = \frac{X^*}{X_s} \qquad (6)$$

Note that this is done for every frequency to produce the 64 quantization steps. That is, for a certain frequency band, all X^* from the image blocks are divided by their corresponding X_s to result in a set of quantization steps that should be the same for that single band. However, due to rounding errors, not all of the resulting steps are equal. We suggest determining the most frequent value among the resulting steps as the most probable one and assigning it to be the correct quantization step for that frequency band.

Table I shows the sample results for the difference between the estimated Q table and the original table for two quality factors. The X's mark undetermined coefficients. The

TABLE I. DIFFERENCE BETWEEN ESTIMATED AND ORIGINAL Q.

QF = 75								QF = 80							
4	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	X	0	0	0	0	0	0	0	X
0	0	0	0	0	0	X	X	0	0	0	0	0	0	X	X
0	0	0	0	X	X	X	X	0	0	0	0	X	X	X	X

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

estimation is slightly better than that of the maximum peak approach for AC coefficients in [15].

The estimated table is then used to verify the authenticity of the image by computing a distortion measure and then comparing it to a preset threshold, as was shown in Figure 1. In our experiments for forgery detection, we used two distortion measures. An average distortion measure for classifying test images can be calculated as a function of the remainders of DCT coefficients with respect to the original *Q* matrix:

$$B_1 = \sum_{i=1}^{8}\sum_{j=1}^{8} \mathrm{mod}\big(D(i,j), Q(i,j)\big) \qquad (7)$$

where *D(i,j)* and *Q(i,j)* are the DCT coefficient and the corresponding quantization table entry at position (*i,j*), respectively. Large values of this measure indicate that a particular block of the image is very different from the one that is expected and, hence is likely to belong to a forged image. Averaged over the entire image, this measure can be used for making a decision about authenticity of the image.

Usually JPEG compression introduces blocking artifacts. Manufacturers of digital cameras and image processing software typically use different JPEG quantization table to balance compression ratio and image quality. Such differences will also cause different blocking artifacts in the images acquired. When creating a digital forgery, the resulted tampered image may inherit different kind of compression artifacts from different sources. These inconsistencies, if detected, could be used to check image integrity. Besides, blocking artifacts of the affected blocks will change a lot by tampering operations such as image splicing, resampling, and local object operation such as skin optimization. Therefore, the blocking artifact inconsistencies found in a given image may tell the history that the image has been undergone. We use the BA measure proposed in [1] as the other distortion measure for classifying test images:

$$B_2(n) = \sum_{i=1}^{8}\sum_{j=1}^{8} \left| D(i,j) - Q(i,j)\, round\left(\frac{D(i,j)}{Q(i,j)}\right)\right| \qquad (8)$$

where *B(n)* is the estimated blocking artifact for testing block *n*, *D(i,j)* and *Q(i,j)* are the same as in (7).

Fig. 4 shows the results of applying these measures to detect possible composites. Normally dark parts of the distortion image denote low distortion, whereas brighter parts indicate high distortion values. The highest consistent values correspond to the pasted part and hence mark the forged area. For illustration purposes, *inverted* images of the distortion measures for the composite images are shown in Figure 4(d) through (g). Hence, black (inverted white) parts indicate high distortion values and mark the inserted parts. Apparently as quality factor increases, detection performance increases and false alarms decrease. This behavior as expected is similar to that of maximum peak method in [15]. However, we observe better clustering of the foreign part and less false alarms in the maximum peak method that in this method.

III. Experimental Results and Discussion

A. Accuracy Estimation

We created a dataset of image to serve as our test data. The set consisted of 550 uncompressed images collected from different sources (more than five camera models), in addition to some from the public domain Uncompressed Color Image Database (UCID), which provides a benchmark for image processing analysis [16]. For color images, only the luminance plane is investigated at this stage. Each of these images was compressed with different standard quality factors, [50, 55, 60, 65, 70, 75, 80, 85, and 90]. This yielded 550×9 = 4,950 *untouched* images. For each quality factor group in the untouched JPEG set, the luminance channel of each image was divided into 8×8 blocks and the block DCT was applied to reconstruct the dequantized coefficients. Then for each frequency band, all dequantized coefficients were collected and stored in an array while on the other hand, their compressed version were dumped from the image file and stored in a corresponding array. Zero entries were removed from both sets to avoid division by zeros. The next step was to apply (6) and divide the dequantized coefficients over their dumped values. The resulting set of estimated quantization step was rounded and the most frequent value was selected as the correct step for that frequency band. This was repeated for all 64 frequencies to construct the 8×8 luminance quantization table for the image. The resulting quantization table was compared to the image's known table and the percentage of correctly estimated coefficients was recorded. Also, the estimated table was used in equations (7) and (8) to determine the image's average distortion and blocking artifact measures, respectively. These values were recorded and used later to set a threshold value for distinguishing forgeries from untouched.

The above procedure was applied to all images in the dataset. Table II shows the numerical results where we can observe the improvement in performance over the maximum peak method especially for high frequencies. Notice that for QF = 95 and 100, the percentage of correct estimation was 98% and 100% respectively, meaning that the method can estimate small quantization steps in oppose to the maximum peak method.

Maximum Likelihood methods for estimating *Q* tables [8-9], tend to search for all possible *Q(i,j)* for each DCT coefficient over the whole image which can be computationally exhaustive. Furthermore, they can only detect standard compression factors since they re-compress the image by a sequence of preset quality factors. This can also be

TABLE II. Percentage of Correctly Estimated Coefficients for Severla QFs

QF	50	55	60	65	70	75	80	85	90	95	100
Max. Peak	66.9	69.2	72.0	74.2	76.9	79.4	82.3	85.5	88.2	66.33	52.71
Streamed Coeff.	87.94	89.16	90.37	91.37	92.36	93.24	94.11	95.66	97.21	98.61	100

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(a) Original with QF = 80.

(b) Original with QF = 70.

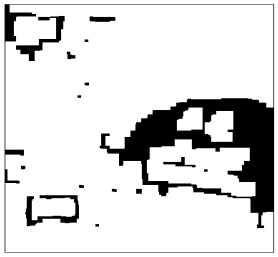
(c) Composite image.

(d) QF = 60

101.255

59.7697

(e) QF = 70



96.2043

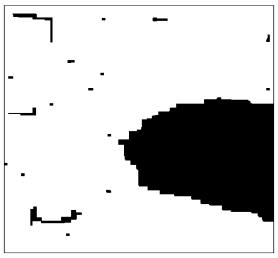
57.7991

(f) QF = 80

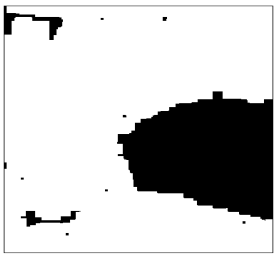
64.5664

49.4292

(g) QF = 90



58.56



43.43

Figure 4. Two test images (a) and (b) used to produce a composite image (c). For each QF (d) through (g), the left column figures represents the average distortion measure while the right column figures represents the blocking artifact measure for the image in (c).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

a time consuming process. Other methods [1, 11] estimate the first few (often first 3×3) low frequency coefficients and then search through lookup tables for matching standard matrices**.** Tables III and IV show the estimation time and accuracy of the proposed streamed coefficients method against the MLE method, power spectrum method, and the maximum peak method, for different quality factors averaged over 500 test images of size 640×480 from the UCID. *Notice that the comparison is based on the estimation of only the first nice AC coefficients, as the two other methods fail to generate estimations for high frequency coefficients.* Notice also that the streamed coefficient method correctly estimated all nine coefficients for all quality factors while requiring the least time.

B. Forgery Detection

To create the image set used for forgery testing, we selected 500 images from the untouched image set. Each of these images was processed in a way and saved with different quality factors. More specifically, each image was subjected to four kinds of common forgeries; cropping, rotation, composition, and brightness changes. Cropping forgeries were done by deleting some columns and rows from the original image to simulate cropping from the left, top, right, and bottom. For rotation forgeries, an image was rotated by 270^{o}. Copy-paste forgeries were done by copying a block of pixels randomly from an arbitrary image and then placing it in the original image. Random values were added to every pixel of the image to simulate brightness change. The resulting fake images were then saved with the following quality factors [60, 70, 80, and 90]. Repeating this for all selected images produced total of (500×4) × 4 = 8,000 images. Next, the quantization table for each of these images was estimated as before and used to calculate the image's average distortion (7), and the blocking artifact, (8), measures, respectively.

Accordingly, the scattered dots in Fig. 5(a) and (b) show the values of the average distortion measure and BAM for the 500 untouched images (averaged over all quality factors for each image) while the cross marks show the average distortion values for the 500 images from the forged dataset. Empirically, we selected thresholds τ = 55 and 35 that corresponded to FPR of 9% and 3% for average distortion measure and BAM respectively. The horizontal lines mark the selected values.

On the other hand, Fig. 6 shows the false negative rate FNR for the different forgeries at different quality factors. The solid line represents the FNR of the average distortion measure, while the dashed line is for the blocking artifact measure. Each line is labeled with the average FNR over all images. Notice the drop in error rates for streamed coefficient method than that of maximum peak method. This is expected since the experiments showed the improved performance of the former method. Notice also that the cropped and composite image sets recorded a zero false negative with BAM. This means that all images in these sets were successfully classified as a forgery. Hence, again the BAM proves to be more sensitive to the types of forgeries especially those that destroy the JPEG grid. Table V summarizes the error rates recorded for the different forgeries.

TABLE III. AVERAGE ESTIMATION ACCURACY (FIRST 3×3) FOR DIFFERENT METHODS

QF / Method	50	60	70	80	90	100
MLE	71.12	85.75	96.25	96.34	80.50	80.3
Power Spectrum	65.37	68.84	75.75	90.12	84.75	84.29
Maximum Peak	96.04	97.69	97.33	91.89	73.33	65.89
Streamed Coeff.	100	100	100	100	100	100

TABLE IV. AVERAGE ESTIMATION TIME (FIRST 3×3) FOR DIFFERENT METHODS.

QF / Method	50	60	70	80	90
MLE	22.29	22.35	22.31	22.26	22.21
Power Spectrum	11.37	11.26	10.82	10.82	11.27
Maximum Peak	11.27	11.29	11.30	11.30	11.30
Streamed Coeff.	0.9336	0.9336	0.9336	0.9336	0.9336

TABLE V. ERROR RATES FOR DIFFERENT TYPES OF IMAGE MANIPULATIONS.

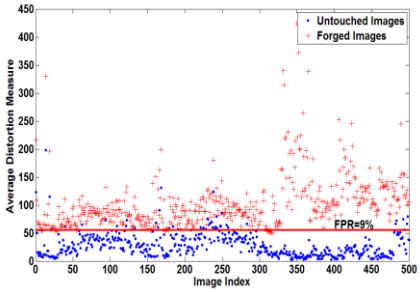
Distortion Measure	Original	Cropp.	Rotation	Compositing	Bright.
Average	9.0%	6.85%	6.5%	6.2%	4.65%
BAM	3.0%	0.0%	4.9%	0.0%	0.55%

IV. DISCUSSION AND CONCLUSIONS

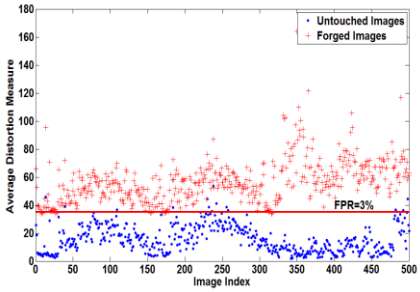
In this paper we have proposed a method for estimating quantization steps based on dumped DCT coefficients from the image file. We have concluded the relation between the constructed dequantized DCT coefficients and their streamed compressed version. We have also verified that while ignoring rounding errors we still can achieve high estimation accuracy that outperformed maximum peak method and two selected methods. Furthermore, we have showed how this method compensates the weak performance for the maximum peak method for high quality factors. We have recorded an accuracy of 98% to 100% for *QF*>90 using the streamed coefficients method.

Through practical experiments we have found that the maximum peak method performs well; by computing a histogram once for each DCT coefficient, quantization steps can be correctly determined even for most high frequencies and hence eliminate further matching or statistical modeling. Naturally this affects execution time (maximum of 60 seconds for a 640×480 image) since we have to process all 64 entries. On the other hand, we have found that the MLE method and power spectrum method outperformed maximum peak method in estimating quantization steps for high qualities. However, for the first 9 AC coefficients, MLE required double the time, and the average time in seconds for the other two methods was found to be very close with an accuracy of 77% for power spectrum as opposed to 91% for maximum peak. Hence, there's trade-off between achieving high accuracy while eliminating the need for lookup tables, and achieving less

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

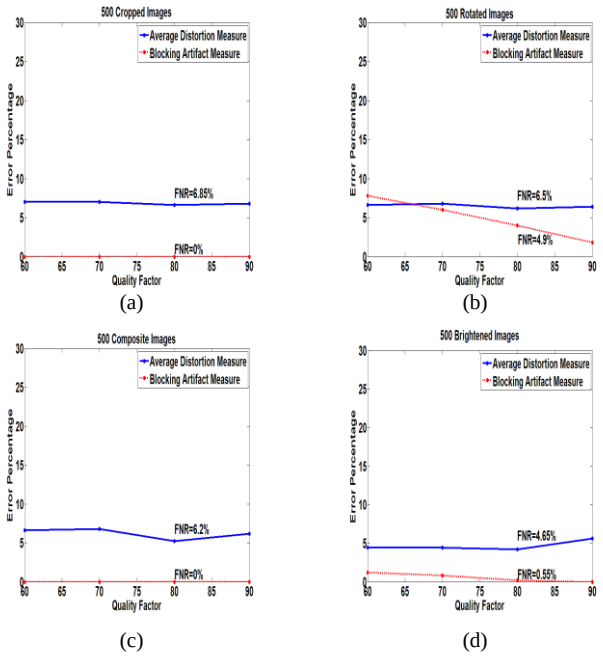


(a) Average distortion measure.



(b) Blocking artifact measure.

Figure 5 Distortion measures for untouched and tampered JPEG images.



(a) (b) (c) (d)

Figure 6 FNR for average distortion measure and blocking artifact measure for (a) cropped (b) rotated (c) composites and (d) rotated JPEG images.

execution time. Nevertheless, we have shown that the proposed streamed coefficients method performed the best with a 100% correct estimation for the first 3×3 AC coefficients for all quality factors with the least execution time.

In addition, we have investigated the use of the estimated quantization tables in verifying the authenticity of images using distortion measures with four common forgery methods. Generally, the performance of the two measures was found to be relatively close for brightened and rotated images. However, BAM was found to be more sensitive to cropping and compositing since it works on the JPEG's grid. Rotation and brightness manipulates were the highest in error rates. They are the most likely to go undetected as they leave the grid intact. On the other hand, streamed coefficients method again outperformed maximum peak method in forgery detection especially with the BAM. As it recorded a zero false negative rate for cropped and composite images.

REFERENCES

[1] Ye S., Sun Q., Chang E.-C., "Detection Digital Image Forgeries by Measuring Inconsistencies in Blocking Artifacts", in *Proc. IEEE Int. Conf. Multimed. and Expo.*, July, 2007, pp. 12-15.

[2] J. Fridrich and J. Lukas, "Estimation of Primary Quantization Matrix in Double Compressed JPEG Images", In Digital Forensic Research Workshop, 2003.

[3] T. Pevný and J. Fridrich, "Estimation of Primary Quantization Matrix for Steganalysis of Double-Compressed JPEG Images", P*roc. SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, pp. 11-1-11-13, San Jose, CA, January 28-31, 2008.

[4] J. He, et al., "Detecting Doctored JPEG Images via DCT Coefficient Analysis", Lecture Notes in Computer. Science, Springer Berlin, Vol. 3953, pp. 423-435, 2006.

[5] Popescu A., Farid H., "Exposing Digital Forgeries by Detecting Traces of Resampling", *IEEE Trans. Signal Process*, 53(2): 758–767, 2005.

[6] Fridrich J., Soukal D., Lukas J., "Detection of Copy-Move Forgery in Digital Images", *Proc. Digit. Forensic Res. Workshop*, August 2003.

[7] Ng T.-T., Chang S.-F., Sun Q., "Blind Detection of Photomontage Using Higher Order Statistics," in *Proc. IEEE Int. Symp. Circuits and Syst*, vol. 5, May, 2004, pp. 688-691.

[8] Fan Z., de Queiroz R. L., "Maximum Likelihood Estimation of JPEG Quantization Table in The Identification of Bitmap Compression History", in *Proc. Int. Conf. Image Process. '00*, 10-13 Sept. 2000, 1: 948–951.

[9] Fan Z., de Queiroz R. L., "Identification of Bitmap Compression History: JPEG Detection and Quantizer Estimation", in *IEEE Trans. Image Process.*, 12(2): 230–235, February 2003.

[10] Fu D., Shi Y.Q., Su W., "A Generalized Benford's Law for JPEG Coefficients and its Applications in Image Forensics", *in Proc. SPIE Secur., Steganography, and Watermarking of Multimed. Contents IX*, vol. 6505, pp. 1L1-1L11, 2007.

[11] Swaminathan A., Wu M., Ray Liu K. J., "Digital Image Forensics via Intrinsic Fingerprints", *IEEE Trans. Inf. Forensics Secur.*, 3(1): 101-117, March 2008.

[12] Farid H., "Digital Image Ballistics from JPEG Quantization," *Department of Computer Science, Dartmouth College, Technical. Rep*ort TR2006-583, 2006.

[13] Farid H., "Digital Ballistics from JPEG Quantization: A Follow-up Study," *Department of Computer Science, Dartmouth College, Technical. Rep*ort TR2008-638, 2008.

[14] Farid H., "Exposing Digital Forgeries from JPEG Ghosts," in *IEEE Trans. Inf. Forensics Secur.*, 4(1): 154-160, 2009.

[15] Hamdy S., El-Messiry H., Roushdy M. I., Kahlifa M. E, "Quantization Table Estimation in JPEG Images", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 1, No. 6, Dec 2010.

[16] Schaefer G., Stich M., "UCID – An Uncompressed Color Image Database", *School of Computing and Mathematics, Technical. Report*, Nottingham Trent University, U.K., 2003.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

GPS L2C Signal Acquisition Algorithms for Resource-Limited Applications in Challenging Environments

Nesreen I Ziedan
Computer and Systems Engineering Department
Faculty of Engineering, Zagazig University
Zagazig, Egypt
ziedan@ieee.org

***Abstract*—Many emerging indoor and wireless applications require the positioning capabilities of GPS. GPS signals, however, suffer from attenuations when they penetrate natural or man-made obstacles. Conventional GPS receivers are designed to detect signals when they have a clear view of the sky, but they fail to detect weak signals. This paper introduces novel algorithms to detect the new GPS L2C civilian signal in challenging environments. The signal structure is utilized in the design to achieve high sensitivity with reduced processing and memory requirements to accommodate the capabilities of resource-limited applications, like wireless devices.**

The L2C signal consists of a medium length data-modulated code (CM) and a long length dataless code (CL). The CM code is acquired using long coherent and incoherent integrations to increase the acquisition sensitivity. The correlation is calculated in the frequency domain using an FFT-based approach. A bit synchronization method is implemented to avoid acquisition degradation due to correlating over the unknown bit boundaries. The carrier parameters are refined using a Viterbi-based algorithm. The CL code is acquired by searching only a small number of delays, using a circular correlation based approach. The algorithms' computational complexities are analyzed. The performances are demonstrated using simulated L2C GPS signals with carrier to noise ratio down to 10 dB-Hz, and TCXO clocks.

***Index Terms*—GPS, L2C, Acquisition, Weak Signal, Indoor, Viterbi**

I. Introduction

The Block IIR-M GPS satellite series started the transmission of a new and more robust civil signal on the L2 carrier frequency- the signal is known as L2C. The first satellite in the series was launched in September 2005, and by August 2009, the eighth and final IIR-M satellite was launched. The L2C signal [1] [2] has different structure and enhanced properties over the GPS L1 C/A signal. The L2C codes and the C/A code have a chipping rate of 1.023 MHz. The C/A signal is modulated by a 1023-chip code, and a 50 Hz data message. The code repeats every 1 ms, and each data bit has exactly 20 codes. While the L2C signal consists of two codes, CM and CL, that are multiplexed chip-by-chip, i.e. a chip of the CM code is transmitted followed by a chip of the CL code. The chipping rate of each code is 511.5 KHz. The CM code has a length of 10230 chips; it repeats every 20 ms, and it is modulated by a 50 Hz data message. The data and the CM code are synchronized such that each data bit has exactly one code. The CL code is 75 times longer than the CM code (767,250 chips), and it is data-less. Performance evaluations for the L2C signal were presented in [3] [4].

GPS signals suffer from attenuation if their paths are obstructed by natural or man-made objects- such as trees or buildings. Conventional GPS receivers can detect signals if their carrier to noise ratio, C/N_0, is over 35 dB-Hz, but they fail to detect weaker signals. Special algorithms are needed to acquire and track weak signals. Many devices that are prone to receiving weak signals, like cell phones, have limited resources. So, the processing and memory requirements must be considered when designing such algorithms.

The acquisition goal is to find the visible satellites, the code delay, τ, and the Doppler shift, f_d. A search for a satellite is done by locally generating its code and using it in a 2-dimensional search on τ and f_d. The received signal is correlated with different versions of a code-modulated local signal, each version is compensated by one possible τ-f_d combination. The codes' properties cause the correlated signals to generate a clear peak only if their codes are the same and their code delays and Doppler shifts are close enough. A positive acquisition is concluded if a correlation exceeds a predefined threshold.

The conventional hardware approach [5] [6] searches for a satellite at each possible code delay and Doppler shift sequentially. Circular correlation [7] [8] uses Fast Fourier Transform (FFT) methods. It calculates the correlation at all the delays at once, for each Doppler shift. Double Block Zero Padding (DBZP) [7] [9] [10] calculates the correlations in the frequency domain, and uses only one version of the replica code. It requires less processing, but it suffers from

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

limitations when working with weak signals. This is because it does not consider the Doppler effect on the code length. The Doppler shift changes the speed of the code, so the code length either shrinks or expands based on the Doppler shift's polarity. This effect can be ignored with small integration lengths, but it will cause acquisition failure with long integration lengths. The problem is that correlating a fixed length local code with a changing length received code will cause the delay to continuously change with respect to the local code. As the integration length increases, the signal power will continue to accumulate at different delays. This will prevent the correct delay from accumulating enough power to exceed the acquisition threshold. This limitation is circumvented in a modified version of DBZP, called MDBZP, which was introduced in [11]. The MDBZP divides the whole Doppler range into a small number of ranges. The correlations in each range are calculated using a version of the replica code that is compensated, in length, by the Doppler shift located in the middle of that range.

A joint acquisition algorithm of the CM and CL codes was introduced in [12]. An assisted acquisition of the CL code was presented in [13]. An FFT-based approach was introduced in [14] to acquire the CM and CL codes. A method called XFAST was introduced in [15] to acquire the long P(Y) code of the L1 signal. This method was extended in [16] to acquire the long CL code; the extended method was called hyper-codes. For resource-limited devices, the problem with acquiring weak signals using the CL code is the high processing and memory required to correlate and search the 767,250-chip code.

This paper introduces acquisition and fine acquisition algorithms for the new L2C signal to work under weak signal conditions. The algorithms utilize the L2C signal structure to achieve high sensitivity and reduced processing and memory requirements. Three algorithms are introduced to work sequentially to first acquire the medium-length CM signal, then refine the estimates of the carrier parameters, and then acquire the long-length CL code. A computational complexity analysis for the algorithms is provided.

The acquisition of the CM code is done using a new version of the MDBZP designed to fit the CM code structure and deal with the fact that each CM code is modulated by one data bit, which has an unknown value. The new algorithm- called CM Acquisition and Bit Synchronization (CM-ABS)- implements a bit synchronization method within the acquisition to avoid correlating over bit boundaries. The correlations are calculated in the frequency domain, and the Doppler effect on the code is considered. Long coherent and incoherent integrations are used, without requiring assisting information from outside sources- like wireless networks. The likely data bit combination is estimated over each coherent integration interval, and used to remove the data signs.

The fine acquisition algorithm is based on the Viterbi Algorithm (VA) [17] [18] [19], which is an optimal dynamic programing technique. The new algorithm is called Fine Acquisition VA-based for L2, or FAVA-L2. The CM code duration is 20 ms, so the phase difference between the start and the end of one code could be relatively large. Using 20-ms correlated signals directly will not provide high accuracy estimation for the carrier parameters. This problem is handled in this paper by dividing the code into small length segments, calculating the correlation for each segment separately, and then using the correlated segments to find fine estimates for the carrier parameters.

The acquisition of the long length CL code is done using a minimized search approach, MS-CL. It uses the estimates of the CM-ABS and FAVA-L2 to acquire the CL code by searching only 75 possible delays. A method is introduced to calculate the coherent integration in smaller steps to avoid processing large number of samples at once.

II. Signal Model

The received L2C signal is down converted to an intermediate frequency (IF), f_{IF}, and sampled at a rate of f_s. The signal model for one satellite is

$$r_{L2C}(t_\delta) = A\,\{d(t_{\delta,\tau})\,C_{M0}(t_{\delta,\tau}) + C_{0L}(t_{\delta,\tau})\} \cos\left(\theta_{n_\delta} + \theta_0 + 2\,\pi\,(f_{IF} + f_{d_0})\,t_\delta + \pi\,\alpha\,t_\delta^2\right) + n(t_\delta), \quad (1)$$

where t_δ is the sampling time. $t_{\delta,\tau} = (t_\delta - \tau)\,\{1 + (f_{d_0} + \alpha\,t_\delta/2)/f_{L2}\}$ is the sampling time taking into account the Doppler effect on the code length. τ is the code delay. f_{d_0} is the initial Doppler shift. α is the Doppler rate. f_{L2} is the L2 carrier frequency. A is the signal amplitude, which is normalized to drive the noise variance to 1 as in [5], i.e. $A = \sqrt{4\,C/N_0\,T_s}$, $T_s = 1/f_s$. d is the navigation data. θ_0 is the initial phase. θ_{n_δ} is the phase noise at t_δ; it is composed of the total phase and frequency clock disturbances. n is a white Gaussian noise (WGN) with zero mean and unit variance. The two codes are modeled such that C_{M0} is a chip-by-chip combination of the CM code and zeros, and C_{0L} is a chip-by-chip combination of zeros and the CL code.

III. CM Acquisition and Bit Synchronization (CM-ABS)

The CM-ABS calculates the correlation in the frequency domain. The separation between the code delays is taken as the sampling time. The number of possible code delays is defined as N_τ. The algorithm produces Doppler bins with frequency separation of $f_{res} = 1/T_I$, where T_I is the coherent integration length. The number of Doppler bins, N_{f_d}, depends on the Doppler shift range, $\pm f_{dcov}$, where $N_{f_d} = 2\,f_{dcov}\,T_I$. The values of the Doppler bins, defined as f_{d_v}, can be calculated as

$$f_{d_v} = \left[v - \frac{N_{f_d}}{2} - 1\right] f_{res}, \;\; v = 1, \ldots, N_{f_d}. \quad (2)$$

The samples in each T_I ms, of the received signal and the replica code, are divided into N_{f_d} blocks. The size of each block is $S_{block} = f_s\,T_I/N_{f_d}$ samples.

Coherent and incoherent integrations are used. The coherent integration length, T_I, can be multiple, N_t, of one data bit

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

length, T_{dms}. The incoherent integration accumulates L results of T_I-ms coherent integrations to get the total integration, $L\,T_I$ ms. The coherent integration is obtained by first generating a T_I/N_{f_d}-ms partial coherent integration, and then generating the T_I-ms coherent integration. The partial coherent integrations, at the N_τ possible code delays, are generated in N_{step} steps. In each step, the partial coherent integrations are generated at a number of adjacent code delays equal to one block size, S_{block}. So, $N_{step} = N_\tau / S_{block}$. The arrangement of the blocks of the replica code and the received signal relative to each other determines which S_{block} partial coherent integrations are generated in each step. Each S_{block} partial coherent integrations are generated by applying circular correlation (FFT/IFFT) between each two corresponding blocks of the replica code and the received signal.

Each CM code period has exactly one data bit, which has an unknown value. To avoid correlating over bit boundaries, a bit synchronization method is implemented within the acquisition. The idea of the algorithm is to search for the start of the code in the received signal by fixing the samples of the replica code and moving forward the samples of the received signal, until the start of the received code coincides with the start of the replica code. This is done instead of searching directly for the code delay in the received signal by allowing the correlation to be calculated using samples from two adjacent received codes, which could have data bits with different polarities that will result in correlation loss. The aforementioned idea is implemented by arranging the replica code's blocks to start at the beginning of the code and fixing that arrangement at all the N_{step} steps, and shifting the received signal's blocks in each step. The received signal's blocks are arranged such that the first block contains the code delays at which the partial coherent integrations are generated. After each step, the first block is discarded, the remaining blocks are shifted forward, and an additional block is added at the end of the blocks. $N_{f_d} + N_{steps} - 1$ blocks are needed to find the N_τ partial coherent integrations.

The process of generating N_τ partial coherent integrations is repeated L times, once for each T_I-ms integrations. In each repetition, the last $(N_{step} - 1)$ blocks used in the generation of the previous coherent integrations are the same as the first $(N_{step} - 1)$ blocks used in the generation of the current coherent integrations. This is because the generation of two consecutive coherent integrations, for the same possible delay, should use two consecutive T_I-ms lengths of the signal. Since circular correlation involves calculating FFT for each block of the received signal, then FFT for each of the $(N_{step} - 1)$ overlapping blocks is calculated only once. The algorithm is illustrated in Fig. 1.

The algorithm's implementation details are as follows. The received signal in (1) is converted to baseband to produce

$$r_c(t_\delta) = r(t_\delta)\, \exp\{-j\,2\,\pi\, f_{IF}\, t_\delta\}\,. \tag{3}$$

The whole Doppler range is divided into a small number, N_{range}, of ranges. Define f_{mid_i} as the middle frequency of

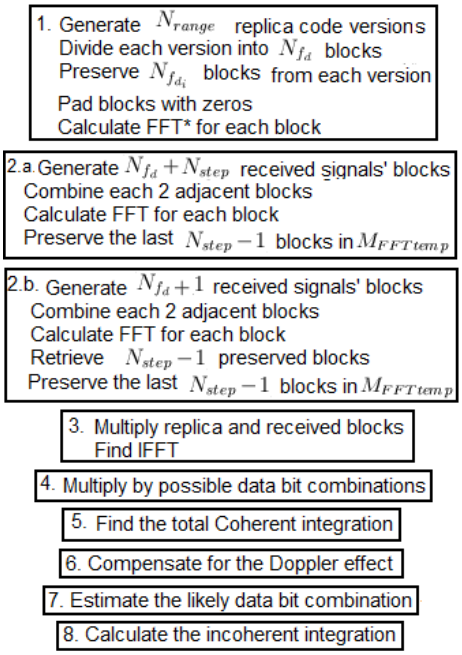


Fig. 1. Illustration of the CM-ABS algorithm.

the i^{th} range, and $N_{f_{d_i}}$ as the number of Doppler bins in the i^{th} range. The indexes of the first Doppler bin and the last Doppler bin, respectively, in each range are

$$\lambda_{s_i} = \sum_{j=1}^{i-1} N_{f_{d_j}} + 1, \tag{4}$$

$$\lambda_{e_i} = \sum_{j=1}^{i} N_{f_{d_i}}. \tag{5}$$

The following items are repeated L times, once for each coherent integration:

1. N_{range} versions of the replica code are generated. Each version is compensated, in length, by one of the f_{mid_i} frequencies. The un-compensated length of each version is T_I ms. The model for each version is

$$C_{LM0_{D\,i}}(t_\delta, f_{mid_i}) = C_{LM0}\left(t_\delta\ \left[1 + \frac{f_{mid_i}}{f_{L2}}\right]\right). \tag{6}$$

Where, $C_{LM0_{D\,i}}$ is the i^{th} replica code version. C_{LM0} is an un-compensated code, which consists of the CM code multiplexed chip-by-chip with zeros. The samples of each version is divided into N_{f_d} blocks, each block has a size of S_{block} samples. From the blocks of the i^{th} version, only $N_{f_{d_i}}$ blocks are preserved, and the others are discarded. The preserved blocks, of the i^{th} version, are those located at offsets from λ_{s_i} to λ_{e_i}. All the preserved blocks are arranged together, where their order is maintained, i.e. blocks coming from the i^{th} replica code are located at offsets from λ_{s_i} to λ_{e_i}. Each block is padded with S_{block} zeros at its end. The complex conjugate of the FFT of each block is calculated. Assume

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

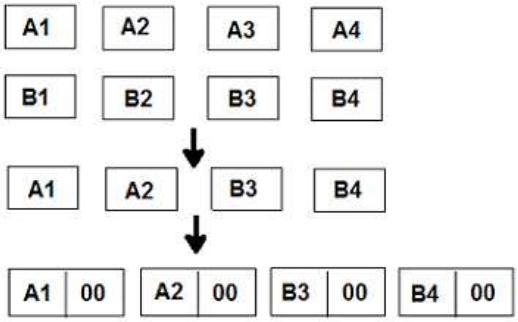


Fig. 2. Example illustrating the local code's blocks arrangement.

a simple example where $N_{range} = 2$ and $N_{f_d} = 4$. Fig. 2 shows the blocks arrangement. The 1^{st} and 2^{nd} replica codes' blocks are defined by (A1, A2, A3, A4) and (B1, B2, B3, B4), respectively. Here, $N_{f_{d_1}} = N_{f_{d_2}} = 2$, and $\lambda_{s_1} = 1$, $\lambda_{e_1} = 2$, $\lambda_{s_2} = 3$, $\lambda_{e_2} = 4$, so A1, A2, B3, and B4 are preserved, while A3, A4, B1, and B2 are discarded. The preserved blocks are then padded with zeros and the processing proceeds.

2. Processing the received samples depends on whether this is the first coherent integration or not. Thus, there are two approaches as follows:

2.a. If this is the first coherent integration, a size of $(N_{f_d} + N_{step})\, S_{block}$ samples of the received signal is divided into $(N_{f_d} + N_{step})$ blocks. Each two adjacent blocks are combined into one block to produce $(N_{f_d} + N_{step} - 1)$ overlapping blocks, with a size of $2\, S_{block}$. FFT is calculated for each block. The last $(N_{step} - 1)$ blocks are preserved in a matrix $M_{FFTtemp}$ to be used in the next coherent integration. Returning to our example where $N_{f_d} = 4$, assume that $N_{step} = 3$. Fig. 3-a illustrates the blocks arrangement. The received samples are divided into 7 blocks, defined as R1 to R7. Each 2 adjacent blocks are combined to form 6 blocks. After FFT is calculated, the 2 blocks marked as (R5, R6) and (R6, R7) are preserved in $M_{FFTtemp}$.

2.b. If this is not the first coherent integration, a size of $(N_{f_d} + 1)\, S_{block}$ samples of the received signal is divided into $(N_{f_d} + 1)$ blocks. Each two adjacent blocks are combined into one block to produce N_{f_d} overlapping blocks. FFT is calculated for each block. The blocks preserved in $M_{FFTtemp}$ are added at the start of the N_{f_d} blocks. The last $(N_{step} - 1)$ blocks are preserved in $M_{FFTtemp}$, overwriting the previous preserved blocks. Returning to our example, Fig. 3-b illustrates the blocks arrangement in the second coherent integration. Only 5 blocks are generated, where the first block will be block R7 from the previous step. Again, each two adjacent blocks are combined to form 4 blocks. After FFT is calculated, the 2 blocks preserved from the previous step, (R5, R6) and (R6, R7), are added at the start of the 4 blocks to complete the 6 blocks needed in the processing of this step. The 2 blocks marked as (R9, R10) and (R10, R11) are preserved in $M_{FFTtemp}$.

3. Items 3.a and 3.b are done to generate the partial coherent integrations for S_{block} delays. They are repeated N_{step} times to calculate the N_τ partial coherent integrations. The index of the first used received signal's block is increased by one in each step to simulate the shifting process, which is explained earlier in this section.

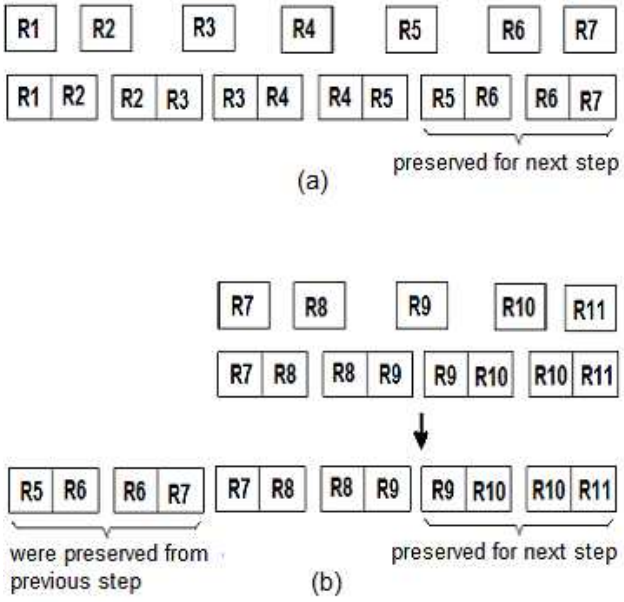


Fig. 3. Example illustrating the received signal's blocks arrangement.

3.a. At the m_r^{th} step, of the N_{step} steps, the blocks at indexes from m_r to $(m_r + N_{f_d} - 1)$, which are generated in item 2, are multiplied block-wise by the blocks generated in item 1. IFFT is calculated for the multiplication result. This produces N_{f_d} blocks of size $2\, S_{block}$ points.

3.b. The first S_{block} points from each produced block are preserved, while the rest are discarded. The preserved points are arranged to form a matrix of size N_{f_d}xS_{block}. The i^{th} row contains the results of the i^{th} block, and the j^{th} column contains the results at index j from each block. This matrix is appended to a matrix, M_c, at row indexes from 1 to N_{f_d} and at column indexes from $[(m_r - 1)\, S_{block} + 1]$ to $[m_r\, S_{block} + 1]$. At the end of the N_{step} steps, M_c will have a size of N_{f_d}xN_τ.

4. Each cell in the M_c matrix is generated using samples of size T_I/N_{f_d} ms, where $T_I = N_t\, T_{dms}$. Each N_{f_d}/N_t rows are generated using samples of size equal to one data bit interval, T_{dms}. To remove the effect of the data bit signs, the M_c matrix is multiplied by the possible data bit combinations. Since the purpose is to have data bits with the same sign over each coherent integration interval, only half the possible data bit combinations are considered, i.e. $2^{N_t - 1}$ possible combinations. This will generate $2^{N_t - 1}$ matrices. Define these matrices as $M_{c_{Ei}}$, where $E = 1, \ldots, 2^{N_t - 1}$.

5. The T_I-ms coherent integrations are found by applying FFT to each column in the $M_{c_{Ei}}$ matrices. Each cell (i, j), of the results, corresponds to the T_I-ms coherent integration at the i^{th} Doppler shift and the j^{th} code delay.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

6. To allow the signal power to accumulate at the correct delay, a compensation is done to account for the difference in the Doppler effect on the code length between a middle range f_{mid_i} and each frequency in that range. Since each row in $M_{c_{Ei}}$ corresponds to a possible Doppler shift, it is circularly shifted by a number of samples, N_{s_v}, which depends on the difference between f_{mid_i} and the frequency value, f_{d_v}, and the coherent integration length T_I and its index, l, within the total incoherent integration. Where

$$N_{s_v} = \text{sign}(f_{d_v} - f_{mid_i})\text{round}\left[(l-1)T_I f_s \frac{|f_{d_v} - f_{mid_i}|}{f_{L1}}\right].$$

7. Only one of the 2^{N_t-1} matrices is preserved and the others are discarded. The preserved matrix corresponds to the likely data bit combination, which is estimated by a method we developed in [11], chapter 3, and is briefly described below.

8. The preserved matrix is added incoherently to the previous total incoherent integration.

After calculating the final total incoherent integration, the cell that contains the maximum power will correspond to the estimated code delay, $\hat{\tau}$, and Doppler shift, $\hat{f}_d$, given that the power exceeds the acquisition threshold. After an acquisition is concluded, the carrier parameters are refined using the fine acquisition algorithm presented in the next section.

Estimation of the likely data bit combination: One or more of the following approaches can be used.

1. Let I_{cE} and Q_{cE} define the real and imaginary parts of the l^{th} coherent integration after the multiplication by a possible data combination, E. Define $P_{l,E}$ as the incoherent integration after including that coherent integration, $P_{l,E}(\tau_u, f_{d_v}) = P_{l-1}(\tau_u, f_{d_v}) + \left[I_{cE}(\tau_u, f_{d_v})^2 + Q_{cE}(\tau_u, f_{d_v})^2\right]_l$. Each $P_{l,E}$ is a matrix of size N_τxN_{f_d}. The matrix that contains the cell that has the maximum power, out of all the matrices, is chosen as P_l, and the others are discarded. The problem with applying this approach directly with weak signals is that the cell that has the maximum power could be a noise cell.

2. The likely data combination is estimated separately for each cell. The 2^{N_t-1} matrices are compared cell wise, i.e. cells at the same index in each matrix are compared. The maximum from each index forms the new integration. At a delay τ_u and Doppler f_{d_v} the new integration is

$$P_l(\tau_u, f_{d_v}) = \max\left\{P_{l1}(\tau_u, f_{d_v}), P_{l2}(\tau_u, f_{d_v}), \ldots, P_{l2^{N_{t_k}-1}}(\tau_u, f_{d_v})\right\}.$$

There will be no loss due to a wrong combination in the correct cell, but the noise power might increase.

3. The likely data bit combination is estimated once for each code delay. The data combination is the one that generates the maximum power among all the Doppler shifts, in the $2^{N_{t_k}-1}$ matrices, at each code delay.

The algorithm can start with the 2^{nd} or the 3^{rd} approach, then switch to the 1^{st} one after several integrations. This is to increase the sensitivity and reduce the degradation due to a wrong data bit combination.

IV. Fine Acquisition Based on The VA (FAVA-L2)

The model for the local signal is

$$IQ_{LM0}(t_\delta) = C_{LM0}\left(t_\delta\left(1 + \frac{\hat{f}_{d_0} + \hat{\alpha}\, t_\delta/2}{f_{L2}}\right)\right) \exp\left\{j\left(\hat{\theta}_0 + 2\pi\left(f_{IF} + \hat{f}_{d_0}\right) t_\delta + \pi\,\hat{\alpha}\, t_\delta^2\right)\right\}, \quad (7)$$

where $\hat{f}_{d_0}$ is the estimated initial Doppler shift, $\hat{\alpha}$ is the estimated Doppler rate, and $\hat{\theta}_0$ is the estimated initial phase. The CM code duration is 20 ms, so the phase difference between the start and the end of one code could be relatively large. Thus, using 20-ms correlated signals in a fine acquisition algorithm will not provide high accuracy estimation for the carrier parameters. This problem is overcome by dividing the code into N_{cFrag} segments, calculating the correlation for each segment separately, and then using the correlated segments to find the fine estimates. Considering the Doppler effect, the estimated code length is defined as T_{d_m} in the m^{th} code period. The integration is calculated over several codes, $N_{d\alpha}$. Hence, there will be $N_{cFrag}\, N_{d\alpha}$ correlated segments. Each correlated segment is expressed as

$$y_i = A\, d_i\, R(\tau_{e_i})\,\text{sinc}\left(\left(f_{e_i} + \alpha_e\frac{T_i}{2}\right) T_i\right) \exp\left\{j\left(\theta_{e_i} + 2\pi\, f_{e_i}\frac{T_i}{2} + 2\pi\,\alpha_e\frac{T_i^2}{6}\right)\right\} + n_{y_i}, \quad (8)$$

where T_i is the segment length. n_{y_i} is a noise term. $R(.)$ is the autocorrelation function. τ_{e_i} is the code delay error. θ_{e_i} and f_{e_i} are the phase and Doppler shift errors, respectively, at the start of the i^{th} segment. α_e is the Doppler rate error. $f_{e_i} = f_{e_{i-1}} + \alpha_e\, T_i + W_{f_d,i}$. $\theta_{e_i} = \theta_{e_{i-1}} + 2\pi\, f_{e_i}\, T_i + W_{\theta,i}$. $W_{\theta,i}$ and $W_{f_d,i}$ are the clock phase and frequency disturbances.

FAVA-L2 works in three stages to find fine estimates for the phase, Doppler shift and rate. First, a Doppler rate error, α_{f_e}, is estimated under the assumption that θ_{e_i} and f_{e_i} are zeros. Second, both α_e and f_{e_0} are estimated, where $\alpha_e = \alpha_{f_e} - (3/2)\,(f_{e_0}/T_\alpha) - E_\alpha$. T_α is the total data intervals used to obtain α_{f_e}, and E_α is an estimation error. Third, the phase error is estimated. This method is similar to a method we introduced in [11], chapter 4, for the C/A L1 signal. This section focuses on describing the modifications and new approaches that are developed for the L2C signal.

α_{f_e} is obtained as follows. Define N_α as the number of possible Doppler rate errors, and α_t as a possible error. The segments that belong to the m^{th} code, defined in (8), are counter rotated by each α_t, and then added, so

$$S_{m,\alpha_t} = \sum_{i=N_{cFrag}\,(m-1)+1}^{N_{cFrag}\,m} y_i \exp\left\{-j\,\pi\,\alpha_t\left(T_{t_i}^2 + \frac{1}{3}T_i^2\right)\right\}, \quad (9)$$

where $T_{t_i} = \sum_{n=1}^{m-1} T_{d_n} + \sum_{k=N_{cFrag}\,(m-1)+1}^{i-1} T_k$. The data bit values are jointly estimated with α_{f_e}. The estimation can be expressed by

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\alpha_{f_e} = \arg \min_{\alpha_t,\, \{\bar{d}_m\}_{m=1}^{N_{d\alpha}}} \sum_{m=1}^{N_{d\alpha}} \left(\hat{d}_m |S_{m,\alpha_t}| - \bar{A}\, \bar{d}_m \right)^2, \quad (10)$$

where $\hat{d}_m$ is an estimated value for the m^{th} bit, which is obtained as the sign of the real part of S_{m,α_t}. $\bar{A}$ is the expected signal level. $\bar{d}_m$ is a possible data bit value (± 1). The estimation process is based on the VA. Define,

$$\Lambda_{N_{d\alpha},\, \alpha_t,\, p} = \sum_{m=1}^{N_{d\alpha}} \left(\hat{d}_m |S_{m,\, \alpha_t}| - \bar{A}\, d_{m,\, p} \right)^2, \quad (11)$$

where p defines a possible data sequence, $p = 1, \ldots, 2^{N_{d\alpha}}$. $d_{m,\, p}$ is the m^{th} data in the p^{th} sequence. Each combination of p and α_t defines a path in a trellis graph. The algorithm operates recursively to find the path that generates the minimum $\Lambda_{N_{d\alpha},\, \alpha_t,\, p}$. Define $\Gamma_{m,\, \alpha_t,\, p_m}$ as the optimal $\Lambda_{N_{d\alpha},\, \alpha_t,\, p}$ in the m^{th} step, for α_t. Since there are N_α of α_t, there are N_α of $\Gamma_{m,\, \alpha_t,\, p_m}$. Since an optimal path consists of optimal sub-paths, it can be shown that

$$\Gamma_{m+1,\alpha_t,p_{m+1}} = \Gamma_{m,\alpha_t,p_m} + \min \left\{ (\hat{d}_{m+1}|S_{m+1,\alpha_t}| - \bar{A})^2, (\hat{d}_{m+1}|S_{m+1,\alpha_t}| + \bar{A})^2 \right\}.$$

After each recursive step, only the N_α paths that correspond to $\Gamma_{m,\, \alpha_t,\, p_m}$ are retained, while the other paths are discarded. At the last step, α_t that corresponds to the minimum cumulative path, among the N_α optimal paths, is taken as the estimated α_{f_e}.

The accuracy of the estimated α_{f_e} depends on the separation between the α_t values. If the range of the possible α_{f_e} is large, then FAVA-L2 is operated iteratively to avoid increasing the processing and memory requirements. The first iteration covers the whole α_{f_e} range, and uses large separation between the α_t's. After each iteration, an estimated α_{f_e} is obtained and used as the middle value of a reduced range. In each iteration, N_α possible values are considered, but the separation between the α_t's is decreased. The process continues until the separation between α_t's is within the desired resolution.

Following the estimation of α_{f_e}, f_{e_0} and α_e are estimated using a similar approach. Define N_{freq} as the number of possible frequencies, and f_t as a possible frequency value. The counter rotation of y_i is

$$S_{m,f_t,E_\alpha} = \sum_{i=N_{cFrag}\,(m-1)+1}^{N_{cFrag}\, m} y_i \exp \left\{ -j2\pi \left(f_t \left(T_{t_i} + \frac{T_i}{2} \right) + \alpha_{f_t,E_\alpha} \left(\frac{T_{t_i}^2}{2} + \frac{T_i^2}{6} \right) \right) \right\},$$

$$\alpha_{f_t,E_\alpha} = \alpha_{f_e} - \frac{3}{2} \frac{f_t}{\sum_{m=1}^{N_{d\alpha}} T_{d_m}} - E_\alpha.$$

E_α models the error in the estimated α_{f_e}. It can be set to a few values, $N_{E\alpha}$, (e.g. 0,1,-1), and thus the number of paths will increase by $N_{E\alpha}$. The estimation is done iteratively, starting with the whole frequency range and with large frequency separation. After each iteration, an estimated f_t is obtained and used as the middle value of a reduced frequency range. After the last iteration, the estimate of $\hat{f}_{e_0}$ is concluded, and the corresponding α_{f_t,E_α} is taken as $\hat{\alpha_e}$. Following that, a similar approach is used to estimate θ_{e_0}. Following the CM-ABS and FAVA-L2, the CL code is acquired using the MS-CL.

V. Minimized Search for The CL Code Acquisition (MS-CL)

The CL and CM codes are synchronized such that the start of the CL code coincides with the start of the CM code. Since the CL code is 75 times longer than the CM code, and the CM code is acquired, then there are only 75 possible code delays to search for the start of the CL code. The local replica is

$$IQ_{L0L}(t_\delta) = C_{L0L} \left(t_\delta \left(1 + \frac{\hat{f}_{d_0} + \hat{\alpha}\, t_\delta/2}{f_{L2}} \right) \right) \exp \left\{ j \left(\hat{\theta}_0 + 2\,\pi \left(f_{IF} + \hat{f}_{d_0} \right) t_\delta + \pi\, \hat{\alpha}\, t_\delta^2 \right) \right\}. \quad (12)$$

Long coherent integration can be used, but the error in the estimated Doppler shift, f_e, will put a limit on its length, T_I, where $T_I \leq 1/f_e$. The coherent integration is generated in steps, with a length of $T_n \leq T_I$, to avoid processing large number of samples at once. The number of steps needed to obtain the total coherent integration is $L_{coh} = \lceil T_I/T_n \rceil$. The coherent integration in the last step could be less than T_n. In addition, L_{CL} coherent integrations are incoherently added to obtain a longer incoherent integration. The total number of steps to get the total incoherent integration is $N_{total} = L_{CL}\, L_{coh}$.

The integrations are calculated for each possible code delay and placed in two vectors, H and P, which hold, respectively, the total coherent and incoherent integrations. U and V are counters initialized to zero. T_k is the total length of the received signal that is used up to the start of the step k, $k = 1, \ldots, N_{total}$; T_k is initialized to zero. The MS-CL works as follows:

1. 75 versions of the CL replica code are used. Each version, C_{L0L_i}, is initialized to start at one of the possible code delays, where $i = 1, \ldots, 75$.

2. The estimates of phase and Doppler shift are propagated to each T_n ms. Define the propagated values as $\hat{\theta}_k$ and $\hat{f}_{d_k}$.

3. The T_n length, taking into account the Doppler effect, is calculated as

$$T_{n_k} = T_n \frac{f_{L2}}{f_{L2} + \hat{f}_{d_k} + \hat{\alpha} T_n/1000/2}. \quad (13)$$

4. 75 versions of the local signal in (12) are generated, each version uses one of the C_{L0L_i} code versions. Each one starts at a time T_k and spans T_{n_k} ms of the CL code. Define each version as IQ_{L0L_i}.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

5. Each IQ_{L0L_i} is correlated with the received signal, starting at T_k. Define each result as Y_i.

6. Y_i is added coherently to the previous total coherent integration, $H(i) = H(i) + Y_i$. The counters are updated, $U = U + 1$, $V = V + 1$.

7. If the desired coherent integration length is reached, i.e. $U = L_{coh}$, then the contents of H are added to the previous total incoherent accumulation, $P(i) = P(i) + \Re\{H(i)\}^2 + \Im\{H(i)\}^2$, where $\Re$ and $\Im$ define the real and imaginary parts, respectively. Following that, H and U are set to zero.

8. If $V < N_{total}$, T_{k+1} is set as $T_k + T_{n_k}$, and then steps (2)–(7) are repeated.

9. If $V = N_{total}$, the CL code delay is concluded from the $P(i)$ that has the maximum power.

To reduce processing as the algorithm progresses, the unlikely code delays can be eliminated. Those will be the delays that generate the minimum $P(i)$. The unlikely code delays can be eliminated every N_{elim} steps, where N_{elim} is set based on the C/N_0, which can be estimated after the fine acquisition. Another method is to eliminate the delays at indexes i_{min} if $P(i_{max})/P(i_{min}) > p_{elim}$. Where, i_{max} is the index of the delay that generates the maximum power at the current step, and p_{elim} is a predefined ratio.

VI. Computational Complexity Analysis

A. CM-ABS

The following are repeated L times to get the total incoherent integration. In item 1, FFT is calculated for N_{f_d} local code's blocks. In item 2.a, FFT is calculated for $(N_{f_d}+N_{step}-1)$ received signal's blocks, while in item 2.b, FFT is calculated for N_{f_d} received signal's blocks. In item 3.a, each two corresponding blocks of the received signal and the local one are multiplied, then IFFT is calculated; this is repeated $N_{step}\,N_{f_d}$ times to get the coherent integration at all the code delays. Each FFT or IFFT operation requires $(2\,S_{block})\log_2(2\,S_{block})$ computations. The number of computations to get the first coherent integration (items 1, 2.a and 3) is

$$\begin{aligned} C_{BP_1} &= [2\,N_{f_d} + N_{step} - 1 + N_{step}\,N_{f_d}] \\ &\quad [2\,S_{block}\log_2(2\,S_{block})] + [2\,S_{block}\,N_{f_d}\,N_{step}]\,. \end{aligned}$$

For the rest of the coherent integrations (items 1, 2.b and 3), this number is

$$\begin{aligned} C_{BP_l} &= [2\,N_{f_d} + N_{step}\,N_{f_d}]\;[2\,S_{block}\log_2(2\,S_{block})] \\ &+ [2\,S_{block}\,N_{f_d}\,N_{step}]\,. \end{aligned}$$

The matrix M_c will have a size of N_{f_d}xN_τ. In item 4, 2^{N_t-1} versions of M_c are generated, each one corresponds to a possible data bit combination. In item 5, FFT is calculated for each column of the 2^{N_t-1} matrices. The number of computations in items 4 and 5 is

$$C_{FFT} = 2^{N_t-1}\,N_\tau\,N_{f_d}\log_2(N_{f_d}). \tag{14}$$

In item 7, if the 1^{st} approach is used to find the likely data combination, each matrix is added incoherently to the previous total incoherent integration. The number of computations is

$$C_{NC} = 2^{N_t-1}\,N_\tau\,N_{f_d}. \tag{15}$$

Only the matrix that corresponds to the likely data combination is kept and the other matrices are discarded. Finding the maximum can be done by one of the sorting methods described in [20], chapter 2. Thus, the number of comparisons needed is

$$C_{CM1} = 2^{N_t-1}\,N_\tau\,N_{f_d} - 1. \tag{16}$$

If the 2^{nd} approach is used to estimate the likely data combination, the number of comparisons is

$$C_{CM2} = 2^{N_t-1}\,N_\tau\,N_{f_d}. \tag{17}$$

In item 8, the real and imaginary parts of each cell are squared and added together. This requires

$$C_{IQ} = 3\,2^{N_t-1}\,N_\tau\,N_{f_d} \text{ computations.} \tag{18}$$

B. FAVA-L2

The operations include generating correlated signals, counter rotating the correlated signals by the possible carrier parameters, and choosing the optimal path. The number of computations can be found directly.

C. MS-CL

The operations include generating 75 versions of the CL local code, multiplying the samples of the received signal and the local one, adding them together to form the coherent integration, and adding the coherent integration to the total incoherent integration. The total number of computations can be shown to be

$$C_{CL} \approx 75\,L_{CL}\,L_{coh}\,[4\,T_n\,f_s + 4]\,. \tag{19}$$

VII. Simulation and Results

The algorithms are demonstrated using simulated GPS L2C signals. The CM code is modulated by ± 1 data with 50 Hz rate and 0.5 probability of data transition. $f_s = 3500$ kHz. $f_{L2} = 1227.6$ MHz. The initial phase is modeled as a uniformly distributed random variable (UDRV) between $(-\pi, \pi)$. A Doppler shift range between $(-5, 5)$ kHz is assumed. The oscillator phase and frequency noises are modeled with normal random walks; the model is similar to the one in [21], and the variances are derived as in [22]. A temperature compensated crystal oscillator (TCXO) is simulated. The values of the phase and frequency random walk intensities are $S_f = 5 \cdot 10^{-21}$ s, and $S_g = 5.9 \cdot 10^{-20}\ s^{-1}$, respectively.

The CM-ABS algorithm is tested using very low C/N_0, 10 and 15 dB-Hz, to demonstrate its ability. A coherent integration length of 80 ms is used. For the 15 dB-Hz signal, a total of 30 incoherent accumulations are calculated. Fig. 4 shows the power versus the code delay. The acquired signal

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

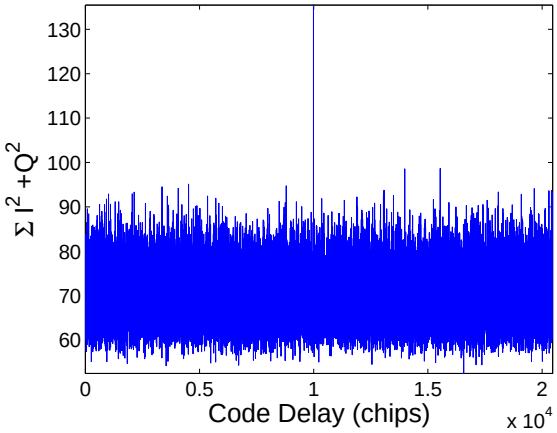


Fig. 4. Power vs. code delay of the acquisition of the CM signal, with $C/N_0 = 15$ dB-Hz, TCXO clock, $T_I = 80$ ms, and 30 incoherent accumulations.

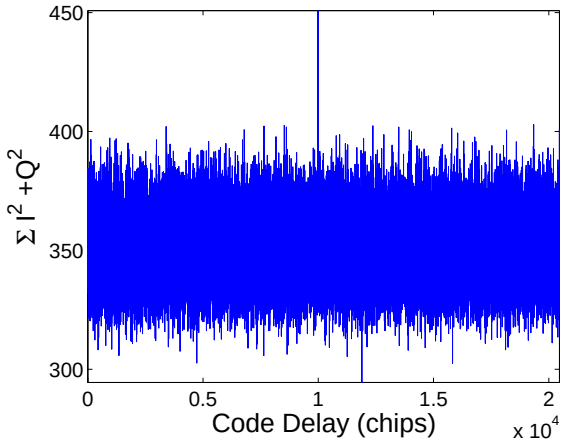


Fig. 5. Power vs. code delay of the acquisition of the CM signal, with $C/N_0 = 10$ dB-Hz, TCXO clock, $T_I = 80$ ms, and 145 incoherent accumulations.

had a power of 135, while the maximum noise power was 100. For the 10 dB-Hz signal, a total of 145 incoherent accumulations are calculated. Fig. 5 shows the power versus the code delay. The acquired signal had a power of 453, while the maximum noise power was 402.

The FAVA-L2 algorithm is tested using C/N_0 between 10 and 24 dB-Hz and TCXO clocks. For this test, the Doppler shift and rate errors are modeled as UDRV's in the ranges of (-50, 50) Hz and (-20, 20) Hz/s, respectively. The algorithm is run for 1000 trials; each trial used 6 seconds of data. The standard deviation (SD) of the estimation error is calculated. Fig. 6 shows the SD of the Doppler rate estimation error versus C/N_0. The SD was about 5.5 Hz/s at 10 dB-Hz, and was about 0.5 Hz/s at 24 dB-Hz. Fig. 7 shows the SD of the Doppler shift estimation error versus C/N_0. The SD was about 8.5 Hz at 10 dB-Hz, and was about 2 Hz at 24 dB-Hz.

The MS-CL algorithm is tested using 15 dB-Hz signal. A

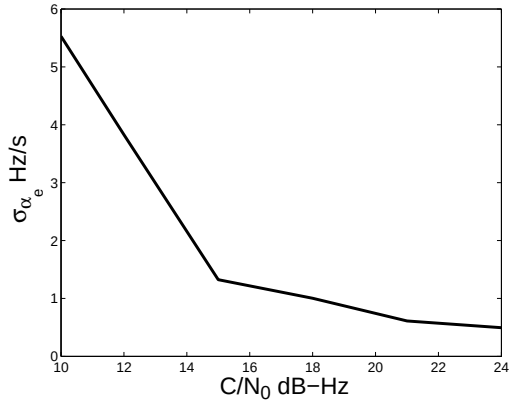


Fig. 6. Standard deviation of Doppler rate estimation error vs. C/N_0 using the FAVA-L2 algorithm, with TCXO clock.

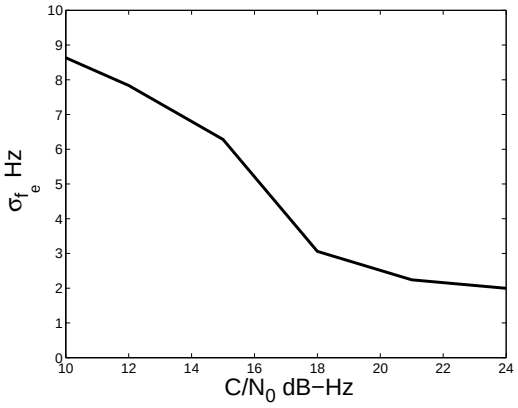


Fig. 7. Standard deviation of Doppler shift estimation error vs. C/N_0 using the FAVA-L2 algorithm, with TCXO clock.

coherent integration of 100 ms is used, with 4 incoherent accumulations. Fig. 8 shows the power versus the 75 possible code delays of the acquisition result. The algorithm correctly estimated the CL code delay at the 31^{st} delay, which generated a power of 52. The maximum noise power was 23.

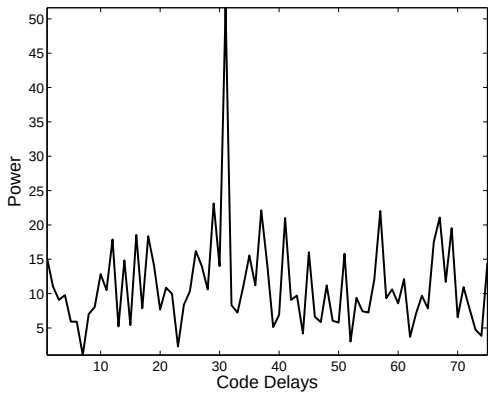


Fig. 8. Power vs. the 75 possible code delays of the acquisition of the CL signal, with $C/N_0 = 15$ dB-Hz, TCXO clock, $T_I = 100$ ms, and a total of 4 incoherent accumulations.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

VIII. Summary and Conclusions

Acquisition and fine acquisition algorithms were presented in this paper for the new L2C GPS signal. The algorithms were designed to work with weak signals, without requiring assisting information from wireless or cellular networks. The paper focused on implementing techniques to increase sensitivity and reduce processing and memory requirements to enable the implementation of the algorithms on devices with limited resources, like wireless devices.

Three algorithms were developed to work sequentially to acquire the CM and CL codes and to provide fine estimates for the carrier parameters. The CM-ABS was designed to acquire the CM code and implemented a bit synchronization method to avoid correlating over bit boundaries. Long coherent and incoherent integrations were used. The Doppler effect on the code length was handled by correctly aligning the coherent integration before adding it to the incoherent accumulation. The FAVA-L2 was designed to provide fine estimates for the phase, Doppler shift and rate. It was based on the optimal Viterbi algorithm. The difference between the total phase at the start and end of the CM code could be relatively large. So, the CM code was divided into small segments, and the correlation for each segment was calculated separately. The carrier parameters were propagated to each segment's time and used to counter rotate each segment's total phase before further processing the segments. The MS-CL was designed to acquire the CL code by searching only 75 possible delays. It used long coherent and incoherent integrations. The integrations can be calculated in smaller steps to avoid exhausting the available resources.

The computational complexities of the algorithms were analyzed. The analysis results can be used to determine the maximum integration lengths that can be used, and the minimum C/N_0 that can be detected based on a device's available resources.

The algorithms were tested using C/N_0 down to 10 dB-Hz and TCXO clocks. The results indicated the ability of the algorithms to work efficiently with such low C/N_0.

References

[1] R. Fontana, W. Cheung, P. Novak, and T. Stansell, "The New L2 Civil Signal," in *ION GPS*, Salt Lake City, UT, September 11–14 2001, pp. 617 –631.

[2] R. Fontana, T. Stansell, and W. Cheung, "The Modernized L2 Civil Signal," *GPS World, 2001*, September 2001.

[3] M. Tran and C. Hegarty, "Performance Evaluations of the New GPS L5 and L2 Civil (L2C) Signals," in *ION NTM*, Anaheim, CA, January 22–24 2003, pp. 521–535.

[4] M. Tran, "Performance Evaluations of the New GPS L5 and L2 Civil (L2C) Signals," *ION Journal of Navigation*, vol. 51, no. 3, pp. 199–212, 2004.

[5] B. Parkinson and J. Spilker, *Global Positioning System: Theory and Applications*. AIAA, 1996.

[6] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jumuna Press, December 1 2001.

[7] J. B. Y. Tsui, *Fundamentals of Global Positioning System Receivers: A Software Approach, Second Edition*. Wiley-Interscience, 2004.

[8] D. J. R. V. Nee and A. J. R. M. Coenen, "New Fast GPS Code-Acquisition Technique Using FFT," *IEEE Electronics Letters*, vol. 27, no. 2, January 17, 1991.

[9] D. M. Lin and J. B. Y. Tsui, "Comparison of Acquisition Methods for Software GPS Receiver," in *ION GPS*, Salt Lake City, UT, September 19–22, 2000, pp. 2385–2390.

[10] D. M. Lin, J. B. Y. Tsui, and D. Howell, "Direct P(Y)- Code Acquisition Algorithm for Software GPS Receivers," in *ION GPS*, Nashville, TN, September 14–17, 1999.

[11] N. I. Ziedan, *GNSS Receivers for Weak Signals*. Artech House, Norwood, MA, July, 2006.

[12] C. Yang, "Joint Acquisition of CM and CL Codes for GPS L2 Civil (L2C) Signals," in *ION AM*, Cambridge, MA, June 27–29 2005, pp. 553–562.

[13] G. S. Ayaz, T. Pany, and B. Eissfeller, "Performance of Assisted Acquisition of the L2CL Code in a Multi-Frequency Software Receiver," in *ION GNSS*, Fort Worth, TX, September 25–28 2007, pp. 1830–1838.

[14] M. L. Psiaki, "FFT-Based Acquisition of GPS L2 Civilian CM and CL Signals," in *ION GNSS 04*, Long Beach, CA, September 21–24 2004, pp. 457–473.

[15] C. Yang, J. Vasquez, and J. Chaffee, "Fast Direct P(Y)-Code Acquisition Using XFAST," in *ION GPS 1999*, Nashville, TN, September 14–17 1999, pp. 317–324.

[16] A. R. A. Moghaddam, R. Watson, G. Lachapelle, and J. Nielsen, "Exploiting the Orthogonality of L2C Code Delays for a Fast Acquisition," in *ION GNSS 06*, Fort Worth, TX, September 26–29 2006, pp. 1233–1241.

[17] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, April 1967.

[18] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[19] R. Riccardo, P. Andreas, and T. Ching-Kae, "Per-survivor processing," *Digital Signal Processing*, vol. 3, no. 3, pp. 175–187, July 1993.

[20] T. Cormen, C. Leisersen, R. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition*. MIT Press, 2001.

[21] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. J Wiley, 1992.

[22] A. J. V. Dierendonck, J. B. McGraw, and R. G. Brown, "Relationship Between Allan Variances and Kalman Filter Parameters," in 16^{th} *PTTI Application and Planning Meeting, NASA Goddard Space Flight Center*, November 27–29, 1984, pp. 273–293.

AUTHOR PROFILE

Nesreen I Ziedan received a Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, USA, in 2004. She also received an M.S. degree in Control and Computer Engineering from Mansoura University in 2000, a Diploma in Computer Networks from the Information Technology Institute (ITI), Cairo, in 1998, and a B.S. degree in Electronics and Communications Engineering from Zagazig University in 1997. Dr. Ziedan is an Assistant Professor at the Computer and Systems Engineering Department, Faculty of Engineering, Zagazig University, Egypt. Dr. Ziedan has several U.S. Patents in GPS receivers design and processing, and she is the author of a book entitled "GNSS Receivers for Weak Signals", which was published by Artech House, Norwood, MA, USA, in 2006.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Method for fingerprint authentication for ATM based banking application

S.Koteswari[#1], Dr.P.John Paul[*2],
V.Pradeep kumar[#1],A.B.S.R.Manohar[#1]
,[#1]*Dept of ECE*
Andhra Pradesh.India.

[*]*Professor, Dept of CSE,*
GATES Engineering College,
Gooty, Ananthapur,
Andhra Pradesh, India.

Abstract: **Fingerprint authentication is widely used in various authentication applications. It is because that fingerprints can achieve the best balance among authentication performance, cost, size of device and ease of use. With identity fraud in our society reaching unprecedented proportions and with an increasing emphasis on the emerging automatic personal identification applications such as biometrics-based verification, especially fingerprint-based identification is preferable as it is used for banking applications. In this paper we are providing authentication using fingerprints of the persons. Here there is two cases train and test. In train case we register the finger print of persons to whom we wish to give authorization .So after register the persons into the data base of the fingerprints .These are changed into templates of predefined .After making Templates the database will be compared with the testing In testing we just make verification after adding the fingerprint of persons. It compares with that template, which are available in database. If it is already in database, it shows matched result else it gives not matched .Finally, we show that the matching performance can be improved by combining the decisions of the matchers based on complementary (minutiae-based and filter based) fingerprint information. The localization of core point represents the most critical step of the whole process. A good matching requires an accurate positioning, so the small errors must also be avoided by usage of complex filtering techniques.**

Keywords-Authentication,Fingerprints,Biometric application,Templates.

I.Introduction:

In today's modern world the automatic authentication of human being is much required in various business applications such as ATM card, Automatic attendance system, Forensic department, Passport verification etc. Various authentication schemes are in use now a day such as signature, fingerprints, retina, DNA etc[1]. But each has some drawbacks either in taking input data or during classification. The devices used to take this data are expensive too. The motivation behind choosing face and finger as biometric is in there ease of collecting input data using very inexpensive devices. The approach is moderately secure for a person cannot change his fingerprints or face. A good recognition system will significantly reduce the manual time required for identification and authentication.

Accurate and automatic identification and authentication of users is a fundamental problem in network environments [2]. Shared secrets such as Personal Identification Numbers or Passwords and key devices like Smart cards are not just enough in some cases. What is needed is something that could verify that you are physically the person you claim to be. The biometrics is enhancing our ability to identify people. And a biometrics system allows the identity of a living person based on a physiological characteristic or a behavioral trait to be verified or recognized automatically. Some of the biometrics used for authentication are Finger Print, Iris, palm print, Hand Signature stroke etc.[3] Among all the biometric techniques, today fingerprints are the most widely used biometric features for personal identification because of their high acceptability, Immutability and individuality. It is a well-known fact that fingerprint is unique to each & every person. These features make the use of fingerprints

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

extremely effective in areas where the provision of a high degree of security is an issue.

The analysis of fingerprints for matching purposes generally requires the comparison of several features of the print pattern. These include patterns, which are aggregate characteristics of ridges, and minutia points, which are unique features found within the patterns. It is also necessary to know the structure and properties of human skin in order to successfully employ some of the imaging technologies.

1.1Patterns:

The three basic patterns of fingerprint ridges are the arch, loop, and whorl. An arch is a pattern where the ridges enter from one side of the finger, rise in the center forming an arc, and then exit the other side of the finger. The loop is a pattern where the ridges enter from one side of a finger, form a curve, and tend to exit from the same side they enter. In the whorl pattern, ridges form circularly around a central point on the finger. Scientists have found that family members often share the same general fingerprint patterns, leading to the belief that these patterns are inherited.

Fig1.1: The arch pattern

Fig1.2: The loop pattern

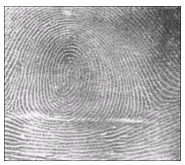


Fig1.3: The whorl pattern

1.2 Minutia features:

The major features of fingerprint ridges are: ridge ending, bifurcation, and short ridge (or dot). The ridge ending is the point at which a ridge terminates. Bifurcations are points at which a single ridge splits into two ridges. Short ridges (or dots) are ridges which are significantly shorter than the average ridge length on the fingerprint. Minutiae and patterns are very important in the analysis of fingerprints since no two fingers have been shown to be identical.

Fig1.4: Ridge ending.

Fig 1.5: Bifurcation.

Fig 1.6: Short Ridge

A smoothly flowing pattern formed by alternating crests (ridges) and troughs (valleys) on the palmar aspect of hand is called a palmprint. Formation of a palmprint depends on the initial conditions of the embryonic mesoderm from which they develop. The pattern on pulp of each terminal phalanx is considered as an individual pattern and is commonly referred to as a *fingerprint*. A fingerprint is believed to be unique to each person (and each finger) 2. Fingerprints of even identical twins are different.

Fingerprints are one of the most mature biometric technologies and are considered legitimate proofs of evidence in courts of law all over the world. Fingerprints are, therefore, used

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

in forensic divisions worldwide for criminal investigations. More recently, an increasing number of civilian and commercial applications are either using or actively considering to use fingerprint-based identification because of a better understanding of fingerprints as well as demonstrated matching performance than any other existing biometric technology.

II.OVERVIEW OF FINGERPRINT

A fingerprint is an impression of the friction ridges of all part of the finger. A friction ridge is a raised portion of the epidermis on the palmar (palm) or digits (fingers and toes) or plantar (sole) skin, consisting of one or more connected ridge units of friction ridge skin. These are sometimes known as "dermal ridges" or "dermal papillae".

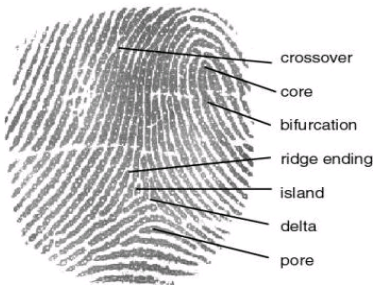


Figure 2.1 A fingerprint image

Fingerprints may be deposited in natural secretions from the eccrine glands present in friction ridge skin (secretions consisting primarily of water) or they may be made by ink or other contaminants transferred from the peaks of friction skin ridges to a relatively smooth surface such as a fingerprint card. The term fingerprint normally refers to impressions transferred from the pad on the last joint of fingers and thumbs, though fingerprint cards also typically record portions of lower joint areas of the fingers (which are also used to make identifications).

A fingerprint is the feature pattern of one finger (Figure 2.1). It is believed with strong evidences that each fingerprint is unique. Each person has his own fingerprints with the permanent uniqueness. So fingerprints have being used for identification and forensic investigation for a long time.

Figure 2.2 A fingerprint image acquired by an Optical Sensor

A fingerprint is composed of many ridges and furrows. These ridges and furrows present good similarities in each small local window, like parallelism and average width.

However, shown by intensive research on fingerprint recognition, fingerprints are not distinguished by their ridges and furrows, but by Minutia, which are some abnormal points on the ridges (Figure 2.3). Among the variety of minutia types reported in literatures, two are mostly significant and in heavy usage: one is called termination, which is the immediate ending of a ridge; the other is called bifurcation, which is the point on the ridge from which two branches derive.

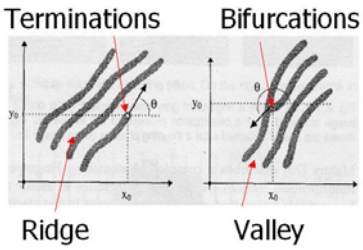


Figure 2.3 Minutia. (Valley is also referred as Furrow,
Termination is also called Ending,
and Bifurcation is also called Branch)

2.1 Authentication vs. authorization:

The problem of authorization is often thought to be identical to that of authentication; many widely adopted standard security protocols, obligatory regulations, and even statutes are based on this assumption. However, more precise usage describes authentication as the process of verifying a claim made by a subject that it should be treated as acting on behalf of a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

given principal (person, computer, smart card etc.), while authorization is the process of verifying that an authenticated subject has the authority to perform a certain operation. Authentication, therefore, must precede authorization. For example, when you show proper identification to a bank teller, you could be authenticated by the teller as acting on behalf of a particular account holder, and you would be authorized to access information about the accounts of that account holder. You would not be authorized to access the accounts of other account holders.

Since authorization cannot occur without authentication, the former term is sometimes used to mean the combination of authentication and authorization.

2.2Authentication vs. Identification:

In the world of virtual identities we find today that many applications and web sites allow users to create virtual identities. Take for example the Second Life world or any chatting forum such as ICQ. The real Identity is hidden and not required. One may actually hold a number of virtual identities. Authentication is still required in order to verify that the virtual identity entering is the original registering identity. The Authentication in this case is of the Login id and not of the person behind it. That requirement poses a problem to most proprietary hardware authentication solutions as they identify the real person behind the virtual identity at delivery.

III. Method for fingerprint authentication

Steps for fingerprint Authentication. figure 3.1 shows the flowchart for finger print authentication

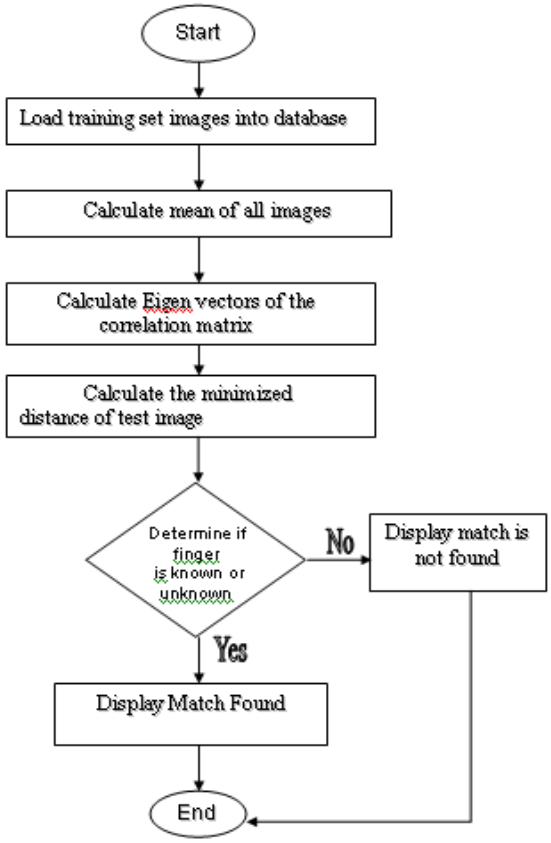


Fig3.1: Flowchart for fingerprint authentication

Step 1: User Registration

In any secure system, to enroll as a legitimate user in a service, a user must beforehand register with the service provider by establishing his/her identity with the provider. For this, the user provides his/her fingerprint through a finger scanner. The finger print image thus obtained undergoes a series of enhancement steps. This is followed by a Finger print hardening protocol with servers to obtain a hardened finger print FP which is stored into the server's database.

Step 2: Fingerprint Enhancement

A fingerprint is made of a series of ridges and furrows on the surface of the finger. The uniqueness of a fingerprint can be determined by the pattern of ridges and furrows. Minutiae points are local ridge characteristics that occur at either a ridge bifurcation or a ridge ending. A ridge termination is defined as the point where a ridge ends abruptly. A ridge bifurcation is defined as the point where a ridge forks or diverges into branch ridges as shown in figure 3.2.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The quality of the ridge structures in a fingerprint image is an important characteristic, as the ridges carry the information of characteristic features required for minutiae extraction.

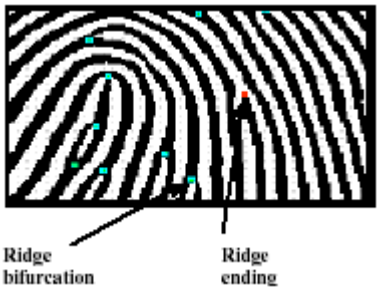


Fig 3.2: Example for ridge bifurcation and ridge ending

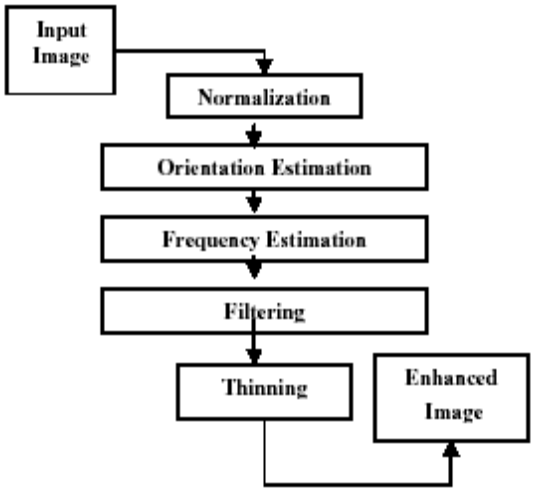


Fig 3.3: Block diagram for fingerprint enhancement

In practice, a fingerprint image may not always be well defined due to elements of noise that corrupt the clarity of the ridge structures. Thus, image enhancement techniques[6] are often employed to reduce the noise and enhance the definition of ridges against valleys. Figure 3.3 illustrates the different steps involved in the development of the Enhancement Finger print. The details of these steps are given in the following subsections.

Step 3: Normalization

Normalization is used to standardize the intensity values in an image by adjusting the range of gray-level values so that it lies within a desired range of values. It does not change the ridge structures in a fingerprint; it is performed to standardize the dynamic levels of variation in gray-level values, which facilitates the processing of subsequent image enhancement stages. Fig. 3.4(a & b) shows the original fingerprint & the results of a normalized fingerprint.

Fig 3.4 (a) Original Image (b) Normalized Image

Step 4: Orientation Estimation:

The orientation field of a fingerprint image defines the local orientation of the ridges contained in the fingerprint (see Fig.3.5). The orientation estimation is a fundamental step in the enhancement process as the subsequent Gabor filtering stage relies on the local orientation in order to effectively enhances the fingerprint image. Fig. 3.6 (a & b) illustrates the results of orientation estimation & smoothed orientation estimation of the fingerprint image.

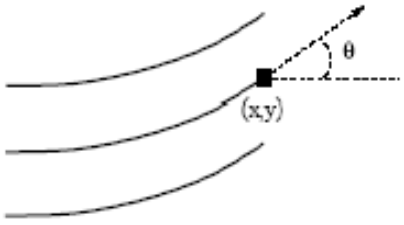


Fig 3.5: The orientation of a ridge pixel in a fingerprint

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

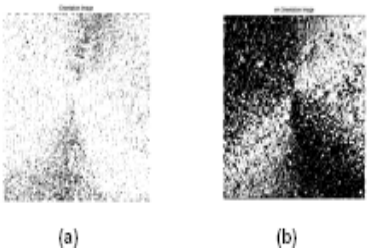


Fig: 3.6 (a)Orientation image (b)Smoothed orientation image

Step 5: Frequency Estimation

In addition to the orientation image, another important parameter that is used in the construction of the Gabor filter is the local ridge frequency. The frequency image represents the local frequency of the ridges in a fingerprint. Fig.3.7, shows the results of the local frequency estimation.

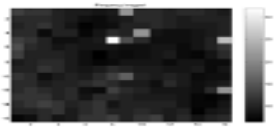


Fig 3.7:Frequency Image

Step 6: Gabor Filtering

Once the ridge orientation and ridge frequency information has been determined, these parameters are used to construct the even-symmetric Gabor filter. Gabor filters are employed because they have frequency-selective and orientation selective properties. These properties allow the filter to be tuned to give maximal response to ridges at a specific orientation and frequency in the fingerprint image. Therefore, a properly tuned Gabor filter can be used to effectively preserve the ridge structures while reducing noise. An even symmetric Gabor filter in the spatial domain is defined as,

$$G(x,y;q,f) = \exp\left\{\frac{1}{2}\left[\frac{x_q^2}{s_x^2} + \frac{y_q^2}{s_y^2}\right]\right\}\cos(2pfx_q),$$

$$x_q = x\cos q + y\sin q$$

$$y_q = -x\sin q + y\cos q$$

where ϴ is the orientation of the Gabor filter, *f* is the frequency of the cosine wave, S_x and S_y are the standard deviations of the Gaussian envelope along the *x* and *y* axes, respectively, and x_q and y_q define the *x* and *y-axes* of the filter coordinate frame, respectively. Fig 3.8 illustrates the results of using gabor filter to a fingerprint image.

Fig 3.8 :Filtered Image

Step 7:Thinning

The final image enhancement step typically performed prior to minutiae extraction is thinning[7]. Thinning is a morphological operation that successively erodes away the foreground pixels until they are one pixel wide. The application of the thinning algorithm to a fingerprint image preserves the connectivity of the ridge structures while forming a skeleton version of the binary image. This skeleton image is then used in the subsequent extraction of minutiae.

The process involving the extraction of minutiae from a skeleton image will be discussed in the next section. Fig. 4.8 illustrates the results of thinning to a fingerprint image.

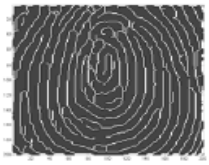


Fig 4.8:Thinned Image

IV. Simulation Results:

Matching algorithms are used to compare previously stored templates of fingerprints against candidate fingerprints for authentication purposes. In order to do this either the original image must be directly compared with the candidate image or certain features must be

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

compared. The training algorithm is shown in figure 4.1.

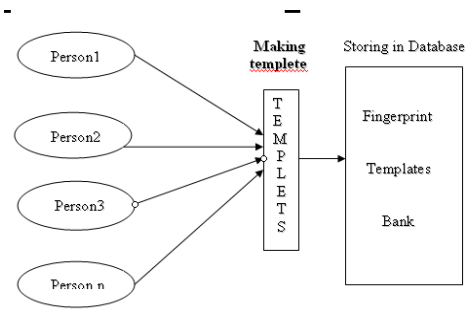


Fig 4.1 : Training algorithm

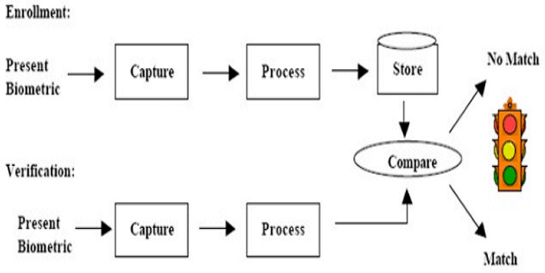


Fig 4.2 : For enrollment fingerprints

In figure 4.2 it indicates the procedure for enrollment of fingerprints. In this paper we are providing authentication using fingerprints of the persons . Here there is two cases test and train. In train case we register the finger print of persons to whom we wish to give authorization .So after register the persons into the database of the fingerprints .These are changed into templates of predefined .After making Templates the database will be compared with the testing. In testing we just make verification after adding the fingerprint of persons. It compares with those templates, which are available in database. If it is already in database, it shows matched result else it gives not matched.The simulation is done in matlab and is designed as shown in figure 4.3 .

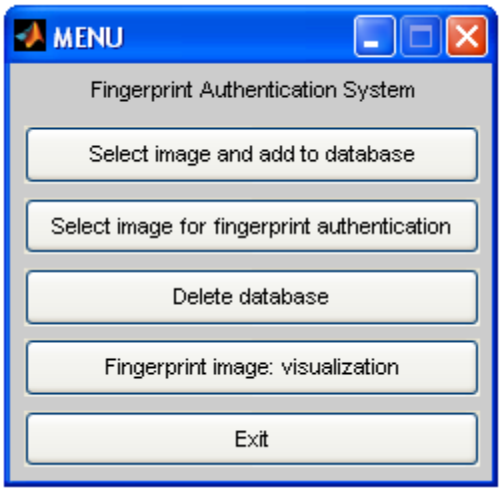


Figure 4.3: simulation result in matlab.

V.CONCLUSION

The salient features of this proposal make it a suitable candidate for number of practical applications like Biometric ATMs and in future, Biometric online web applications etc. Compared with previous solutions, our system possesses many advantages, such as the secure against dictionary attack, avoidance of PKI, and high efficiency in terms of both computation and communications. In this system, we have reused ideas in the areas of image processing technique to extract the minutiae from biometric image. Therefore it can be directly applied to fortify existing standard single-server biometric based security applications.

VI.REFERENCES

[1] W. Ford and B S. Kaliski Jr., "Server-Assisted Generation of a Strong Secret from a Password," Proc. IEEE Ninth Int'l Workshop Enabling Technologies, 2000.
[2] M.Bellare, D. Pointcheval, and P. Rogaway, "Authenticated Key Exchange Secure Against Dictionary Attacks," Advances in Cryptology Eurocrypt '00, 2000 pp.
[3] J. Brainard, A. Juels, B. Kaliski, and M. Szydlo, "A New Two – Server Approach for Authentication with Short Secrets," Proc. USENIX Security Symp., 2003.
[4] Y.J. Yang, F. Bao, and R.H. Deng, "A New Architecture for Authentication and Key Exchange Using Password for Federated Enterprises, " Proc. 20th Int'l Federation for Information Processing Int'l Information Security Conf. (SEC '05), 2005
[5] Y.J. Yang, F. Bao, and R.H. Deng "A Practical Password – Based Two Server authentication and Key Exchange System", IEEE Transactions
on Dependable and Secure Computing , Vol 3,No . 2, April-June 2006
[6] L. Hong, Y. Wan, and A. Jain, "Fingerprint Image Enhancement: Algorithm and Performance Evaluation," IEEE *Trans. Pattern Analysis and Machine* Intelligence, vol. 20, no.8, 1998, pp.777-789.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7] S. Kasaei, M. D., and Boashash, B. Fingerprint feature extraction using block-direction on reconstructed images. In *IEEE region TEN Conf. digital signal Processing applications, TENCON* (December 1997).
[8] D. Boneh,"The Decision Diffie-Hellman Problem," Proc. Third Int" Algorithmic Number Theory Symp.,

S.Koteswari, she is graduated in B.tech from Grad IETE in 2004, Post graduate in M.tech from JNTU Kakinada in 2007 and presently pursuing Ph.D from Rayalaseema University. She has published several papers in National and International conferences and journals She has been guiding students in the area of VLSI and Image processing .Currently she has been working as Associate Professor in Electronics and Communication Engineering Department D.N.R Engineering college Bhimavaram,, Andhra Pradesh, ,India.

Dr. John Paul graduated (B.E., Electronics and Communication Engineering) from Nagarjuna University in the year 1986 with 1st Class. He had completed his post graduation in the year 1994 (M.E., Digital Systems) with first class from Osmania University.

He also did his research in Computer Science (VLSI architecture for image compression) and achieved doctorate in the year 2008 from Hyderabad Central University.He has published several papers in national and international conferences and journals. Also authored books in electronics published by New Age International, New Delhi, Galgotia publications, New Delhi, and guiding students for Ph.D. in the areas of parallel computing, G3 mobile technologies and SS7 of signaling. He worked as professor, dean and Principal of several engineering colleges for 25 years.

V.Pradeep Kumar, he obtained his Bachelor of Technology in Electronics and Communication Engineering, JNTU University, (2004– April). Master of Technology in Electronics and Communication Engineering, JNTU University, (2008– August). (Specialization on VLSI). He is Expertise in dealing B-Tech and M-Tech Engineering Projects on VLSI Stream using both Vhdl & Verilog. He had been working as Assistant Professor in ECE department B.I.E.T College, Hyderabad, Andhra Pradesh, India.

A.B.S.R.Manohar, he is graduated in B.tech from JNTU Kakinada in 2005, Post graduated in M.tech from Satyabhama University ,Chennai. He has been guiding students in the area of VLSI and Image Processing He had been working as Assistant Professor in ECE department, G.V.I.T Engineering College,Bhimavaram, Andhra Pradesh, India.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Empirical Study of Evolution of Decision Making Factors from 1990-2010.

Mohammed Suliman Al-Shakkah*
School of Computing, College of Arts and Sciences
University Utara Malaysia, UUM
06010 UUM-Sintok, Kedah, Malaysia
alshakkah_11@yahoo.com
alshakkah@gmail.com

Wan Rozaini Sheik Osman
School of Computing, College of Arts and Sciences
University Utara Malaysia, UUM
06010 UUM-Sintok, Kedah, Malaysia
rozai174@uum.edu.my

***Abstract*—The intense competition make DM process important for their survival. There are many factors that affect DM in all types of organizations, especially business. In this qualitative study the result has come out with new view for the decision making processing through (observing) analyzing the nine decision making factors from 1990-2010 from 210 papers which were selected randomly from the available resources. Seven partitions were made for the time period of three years and 30 papers for each period. Qualitative method was used here. By analyzing figures and chart with Microsoft excel, the nine decision making factors were categorized into two groups. The main group consists of five factors: time, cost, risk, benefits, and resources. While the second group of the factors consists of four: financial impact, feasibility, intangibles, and ethics. However, time was the most relevant factor at all. More researches in decision making are needed to solve the problems in organizations and in different scopes related to decisions.**

Keywords- Decision making (DM); decision making process (DMP); decision support system (DSS).

I. INTRODUCTION

Decisions affect a lot of life activities and they are needed by many people in different levels [1]. Information System (IS) is an important area, a review in IS research showed its effect on decision making and the success of organizations [7], [8]. In addition to, IS has several subsets such as Decision Support Systems (DSS). A DSS is a computer based system (an application program) capable of analyzing an organizational data and then presents it in a way that helps the decision makers to make business decisions more efficiently and effectively. Besides that, organizations are so dependent on IS, that is urgent attention are focus on those factors that can help decision makers in processing their decisions efficiently and effectively [9].

This importance of decisions gave motivation to see how to improve decision making in organizations. The purpose of this study is to shed a light on what affects decision making process. Studying decision making factors will increase the understanding of this process of making decisions. In this paper, the frequency of decision making factors is counted over a period of twenty years. More clear vision of decision making will presented through answering the following two questions: follow.

- What are the factors that are important in decision making processing which previously?
- What are the relevant factors in decision making for the period 1990-2010?

Before we start discussing these questions, it is good to know that in the perspective of information system management field, the programmers and researchers had created the decision support system (DSS) to help in making decisions without consultant or detailed analysis [2], DSS firstly created to support decision makers in organizations. However, in the large context such as organization, technology would become a good enabler to support distributed decision making [3].

II. DECISION MAKING

A. Decision Making Factors

Many examples of bad decisions cost organizations a lot of money [4]. A suggestion for instructions and steps that improves the quality of decisions, hence results in better decisions. Also [4] asserted nine decision making factors that were presented as: Time, cost, risk, benefits, resources, financial impact, intangibles, ethics and feasibility. For this the researcher reviewed other researches for these factors in the following section.

B. Previous work

In the beginning from the previous factors, it is good to start by time which was intended as time for implementing the alternative and the effect of delay [4]. This factor is very important and is needed in dynamic decision making [10]. In addition, time is so important for managers through their singular decision making, they face unstructured problems which need to be processed quickly [11].

Cost meant to be cost of the alternatives and its suitability to the budget [4]. Other researcher as [12] proposed algorithm to make the optimal decision making with intelligent decision making systems, cost-benefit analysis was used and trials was

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

done to reduce cost with the same benefit. In the same meaning of lowest cost was by [13] in automation 2.0. Also, a case study was applied for the decision support system courses on documentation of the web-based cost estimator for application Al-Sawaf Trading Center [14].

Risk is related to this alternative [4], where risk is inherent in every activity made by the person, and risk insight with to help decision makers in their decision making process [15]. A affect which is as a feeling-state that from good to bad help in decision making for the manager to care with their choices [16]. For the benefit factor which is the profits from implementing this alternative [4], some of the recommendation systems can modalize the customer decision making with high level of decision variable benefits for in the decision making process [17]. Also, using question answering which is related with ontology technique and the data warehousing through application business intelligence bring a lot of benefits for the decision makers [18].

Resources which is for each alternative, the required resources are available [4], In the other hand, using analytical hieratical process (AHP) in decision making process through the available resources help decision makers for better decisions [19]. Also, discussing the key concepts of the IT process management will centralize and control the available resources in organizations [20].

Financial impact which mean the effect of costs with time [4]. In the other hand, financial impact of data accuracy on an inventory system is very important. This will lead through using technology to quantify investment in tracking system and many benefits will be gained in decision making process [21]. Also, some other examples of the computer- based information system as enterprise resource planning (ERP) and supply chain management (SCM) are useful in information technology investment for IT managers to reduce time and cost within processing decisions i.e. which give a strong financial impact for decision makers [22].

Ethics factor is to see if this legal or not [4]. Other researcher revealed the ethical side of using internet technology [23], for human values as ethics, they are increasingly used and still in use as a concept in different fields [24]. Also, the ethical multiplicity for different code of ethic through organizations was discussed [25].

Intangible is for what other unrecognized or sudden variables [4]. In addition, intangible and tangible financial resources operated by organizations are very important [26], for helping decision makers, creating many alternatives can help in processing decisions, even these options related to tangible or intangible resources [27]. Also, enterprise information technology costs a lot of money and risky, so information technology asset for this set of tangible and intangible for operation considered [28].

Feasibility which in the mean those alternatives can be implemented realistically [4]. In addition, there is one method of DSS as multi-alternative decision making properties the alternatives, and the feasibility of applying objective technique in order to maximize numbers of alternatives which help in DMP [29]. Also, the benefit-cost deficit model was proposed to explain and predict barrier removal was feasibility; this will help decision makers in their DMP [30]. To sum up, for the nine factors mentioned it will be worthy if the decision makers in organizations look for in their DMP.

The first question done, now for the second research question: What are the most important factors in decision making for any field? This and all these same meaning questions will be answered in this paper with a qualitative empirical study. The study was carried out on all the available resources to study the decision making factors and how they change with time, from the year 1990 until 2010.

C. Processing the Decision Making

Researchers as [5] studied the old decision making methods. They found that in the old method, the decision making was art of the managers and it requires talents, experiences and intuitions, rather than a systematic method. While, in the modern method, there are four steps in decision making: (1) Define the problem (difficulty or opportunity). (2) Construct a model that describes the real-world problem. (3) Identify the possible solutions to model the problem and evaluate the solutions. (4) Compare, choose and recommend potential solutions to a problem. It has to be ensured that sufficient alternative solutions are considered. Also in this book Simon`s steps were presented in four steps to process decision making as: (1) Intelligence. (2) Design. (3) Choice. (4) Implementation. While, [4] gave five steps of decision making process are stated as: (1) Establish a context for success. (2) Frame the issue properly. (3) Generate alternatives. (4) Evaluate the alternatives. (5) Choose the best alternative.

In addition to, [6] clarified steps to the decision-making process also by other researches were as: (1) Identify the problem or issue. (2) Generate alternatives. (3) Ranking the alternatives and select one of them. (4) Implement the selected alternative. (5) Evaluate the outcomes.

However many researchers call for using the systematic way and they browse different steps, either if it is three, four, or five steps the focus in all is the choosing stage which is the meaning of decision, with this also the need become more and more to understand the important attributes (factors) from the nine attributes mentioned previously in the processing decision making to help all types of decision makers to better decisions. for this paper intend to reveal these important an more interested in factors and how it changes with time, in the next section more details about how the work done.

III. METHODOLOGY

Since the interest is to count each factor is its frequency in each year the qualitative method used in this paper, now the important thing appear how this will be done? The systematic way for this comes in the next sub-sections.

A. Implementation of the Methodology

Here some steps were followed in this study as follows: Firstly in this study papers related for decision making factors were selected randomly from the available resources, after that specify the search (advance search) from the year 1990 until

This work is *sponsored by* University Utara Malaysia

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2010, since technology change faster, the periods were divided to seven periods and every period three years as follows:

First period will be as [1990, 1991, 1992], for the second period will be as [1993, 1994, 1995], for the third period will be as [1996, 1997, 1998], for the fourth period will be as [1999, 2000, 2001], for the fifth period will be as [2002, 2003, 2004], for the sixth period will be as [2005, 2006, 2007], and for the last period will be [2008, 2009, 2010].

Secondly from the related work in section 1.1 the nine factors stated, after that tables prepared and from counting the times for the frequency for each factor, the randomly chosen samples were thirty for each period, data was resulted for each period and the range was from zero to thirty for each factors in every period.

TABLE1. YEARS FOR THE PERIOD : [, ,]

#	Title	Author	Time	Cost	Benefits	Financial l impact	Risk	Resource	Intangible	Ethic	Feasibility
1	…										
2	…										
	…	…	…	…	…	…	…	…	…	…	…
Total											

Thirdly after tabulating data we go for representation the data in an understandable, easy effective way, here we use Microsoft excel to represent data by columns, lines, and sectors here are the results: The data for nine factors and the seven periods were inserted.

In brief all the work in section two was to get the data which is the basic thing needed from the resources for the decision makers to process to support their decisions, after that the analysis by any simple tool can analyze the data which is followed in the next section.

IV. ANALYSIS

Through the descriptive analysis a lot of figures were resulted since the work has seven periods with nine factors; so simple calculation it will be 63 figures if we want to browse at least in two different chart types it will be 126 figures in taking each variable alone, for the beneficial better to compare the factors together to judge which is the more important for this from the initial work some relevant figures will be browse here for the purpose of this work, the comment about the figures will in the next section.

V. RESULTS AND DISCUSSION

As mentioned previously we will browse and comment on the important figures; for that will put it in the following sub-sections:

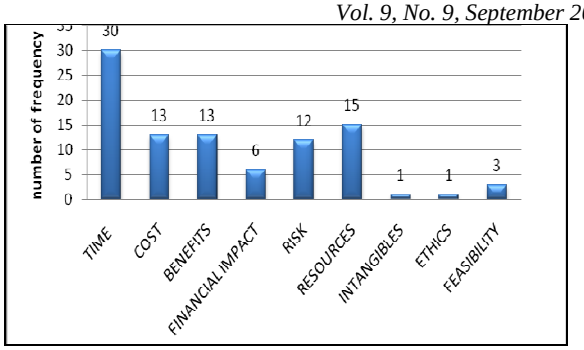


Figure 1.The nine decision making factors in the first period 1990-1992.

Based on Figure 1 the factors for decision making take vary. The number of frequency for time is highest than other factors followed by resources, until lowest number of frequency such as ethics and intangibles. Therefore the first five factors with higher number of frequencies can be considered as: time, cost, benefits, risk, and resources.

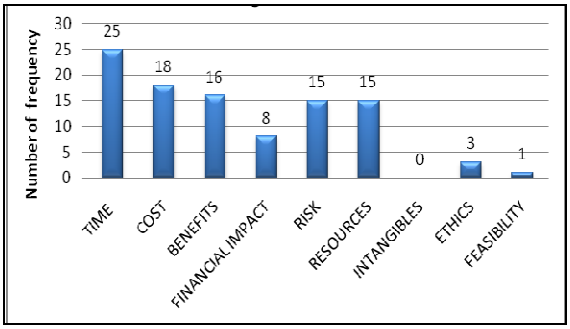


Figure 2. The nine decisions making factors for the year from 1993-1995.

From Figure 2 to rank descending the factors of decision making related to their frequencies it will be as: time, cost, benefits, while risk and resources equal in the fifth position, then the rest of factors.

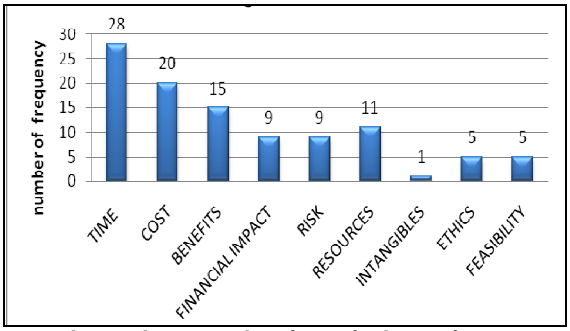


Figure 3. The nine decisions making factors for the year from 1996-1998.

Here in Figure 3 the factors representation obvious as the previous results taking steps shape from time followed by cost then benefits, then the rest of the attributes.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

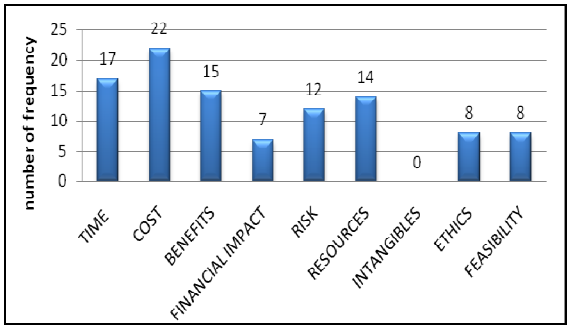


Figure 4: The nine decision making factors for the year from1999-2001.

Based to Figure 4 the time became as second factors while the cost is the first one, in common the same style the first five frequencies still to the following factors: cost, time, benefits, resources, and risk.

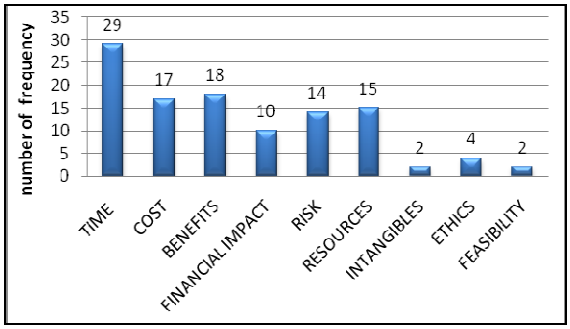


Figure 5. The nine decisions making factors for the years from 2002-2004.

Another support to the near conclusion here by Figure 5 the rank descending for the factors comes out as: time, benefits, cost, resources, risk, financial impact, ethics, feasibility, and intangibles. Also it can be noticed here the same five factors appear again; which is the same results from the following Figure 6 for the period with years from 2005-2007.

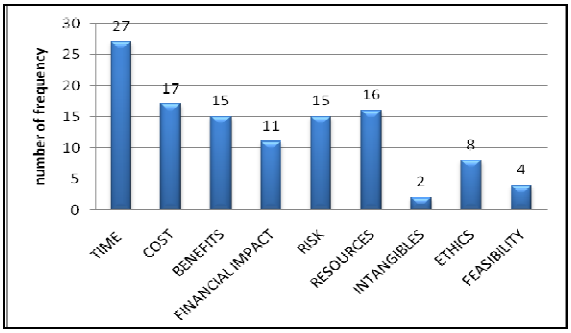


Figure 6. The nine decision making factors for the years from 2005-2007.

For the last period 30 papers will be selected from the available resources for the decision making factors survey.

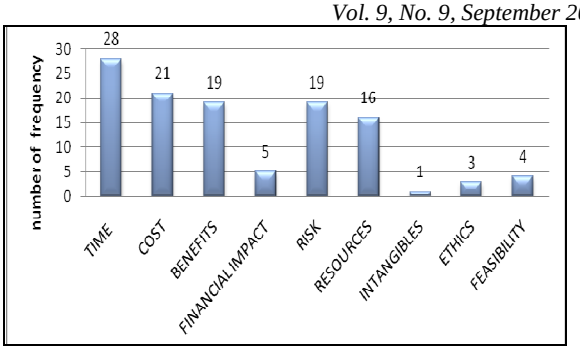


Figure 7. The nine decision making factors for the years 2008-2010.

Descriptive analysis for papers for the years [2008,2009, 2010], in addition to what mentioned previously the same result appeared again one look to the previous figures will conclude the same five factors appear again and this will be a powerful guide to the conclusion in this research paper.

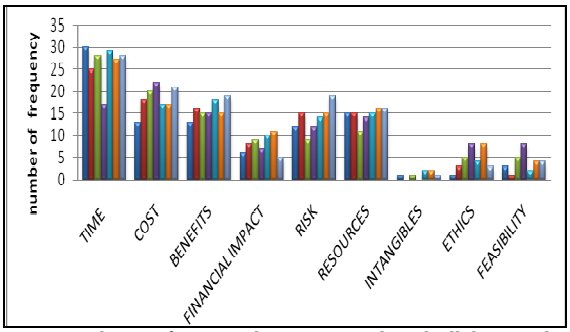


Figure 8: The nine factors in the seven periods with all the periods.

Based on Figure 8 which is considered a comprehensive figure, for each factor seven columns which represent the seven periods for the years from 1990 until 2010, which indicates also to another support for the previous result the descending rank for the factors still grouping the previous five factors as the more interested and wanted to the decision makers from the other factors. Another representation may be preferred to give it in bars some like to see things while comparing in (many views) horizontal view followed here in Figure 9.

More easily view in the following figure to the previous Figure 9 and as a good result the representation in averages for the nine factors for the seven periods as follows in Figure 10.

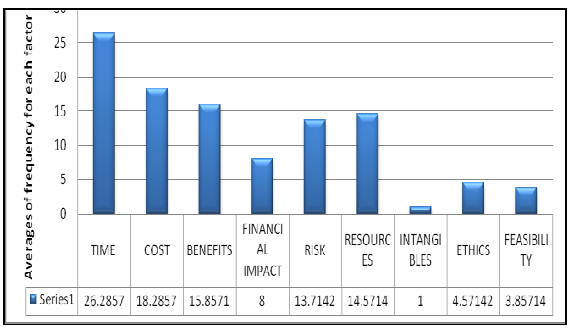


Figure 10. The average of frequency for the nine decision making factors from 1990-2010.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

From the figures presented previously and discussion the factors of decision making can be categorized into two groups: the major group one which consists of five factors: cost, time, risk, resources, and benefits, while the second group consists of four factors: financial impact, feasibility, intangibles, and ethics.

For anyone who will wonder from these five factors which is the more frequently and more redundant with all the years from 1990 until 2010. To give the answer for this wondering we need restart the previous work with partial data from the previous data for the five factors in group one.

However, as mentioned before, no meaning from analyzing the time alone or any other factors, for that the comparison will be between the five factors all in every period from the seven previously mentioned periods then lastly all together.

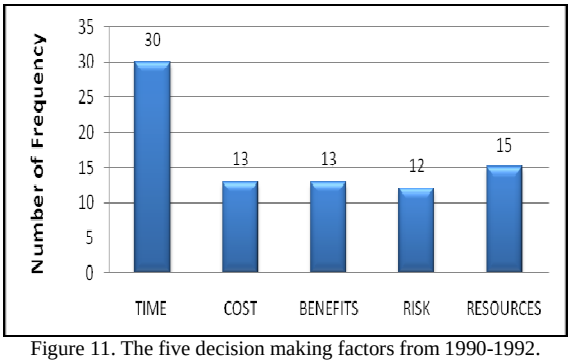


Figure 11. The five decision making factors from 1990-1992.

Based to Figure 11 it represents the first period (1990-1992) clearly time with the most frequency from all the five presented factors, then comes resources, follows by two attributes in the same level: cost and benefits, and at last one is the risk factors.

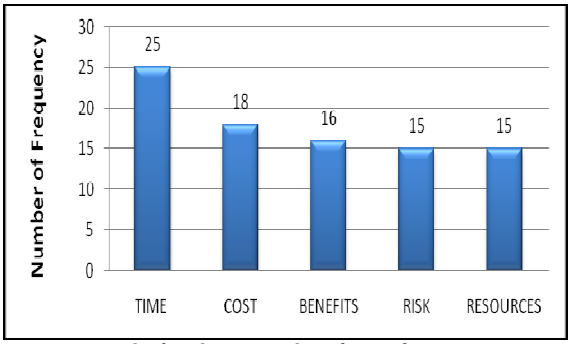


Figure 12. The five decision making factors from 1993-1995.

Here in the second period (1993-1995) based to figure 12, it is easy to notice them as they look like steps, time is the highest, fellows by cost, then benefits, and lastly risk same level as resources.

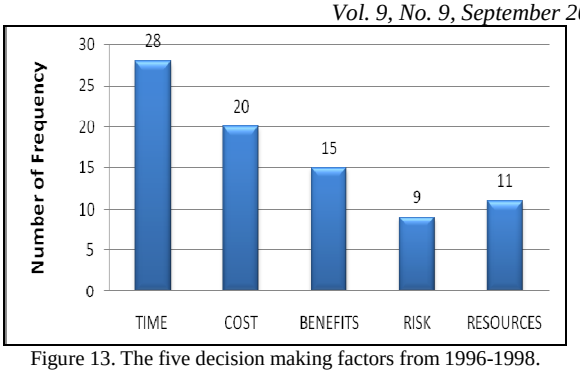


Figure 13. The five decision making factors from 1996-1998.

In the third period (1996-1998) based to Figure 13 time is the highest frequency, the second factors cost, then the third the benefits factors, the resources here is the fourth, and the lowest factor is the risk.

In the following figure has change from the previous style. A fast look for Figure 14 you will see time didn't come in the first stage, so the cost factors come with the highest frequency, but followed by time in the next stage, then the benefits factor after that the resources, and at the end came the risk, see Figure 14.

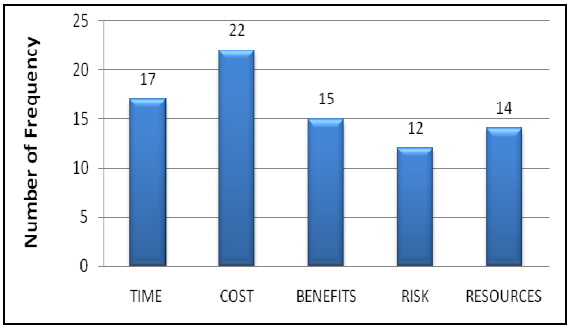


Figure 14. The five decision making factors from 1999-2001.

To reach to meaningful result from the coming figure the focus will be for time to verify is it still the highest, whereas for risk is it still the last one, see Figure 15 the following one, and for the other three factors they varies in different ways.

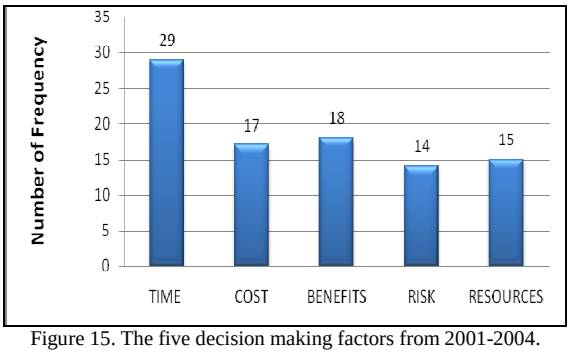


Figure 15. The five decision making factors from 2001-2004.

For the sixth period with the years (2005-2007) the time factor return back to be the highest of all the fifth factors, and the other four factors in different high representation for their

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

frequencies, in the last period the important to track the time factors behavior and ignore the other factors to avoid misleading the issue to come with beneficial result. See the following Figure 16

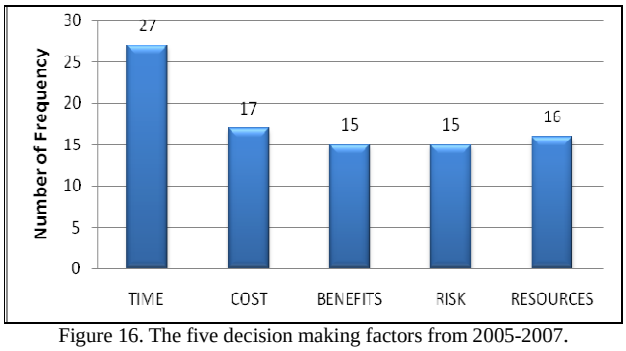


Figure 16. The five decision making factors from 2005-2007.

For the last period for the years 2008-2010, it is obvious the time is the highest column which represents the frequency from the based to the following figure, see Figure 17

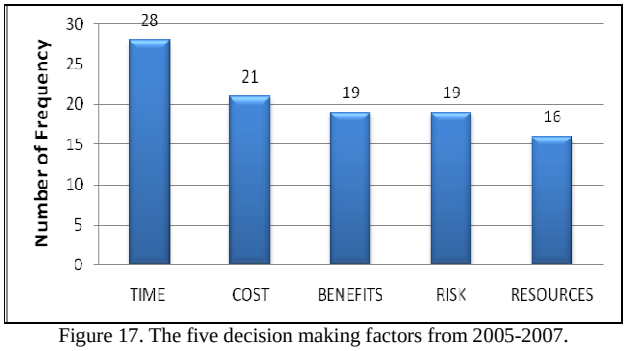


Figure 17. The five decision making factors from 2005-2007.

It is good before coming out with conclusion to have another support, for which is the highest factors or the more relevant one from the five resulted attributes from the initial nine factors , for that the following will be representation to the five factors together in all the seven periods. For this see the following Figure 18

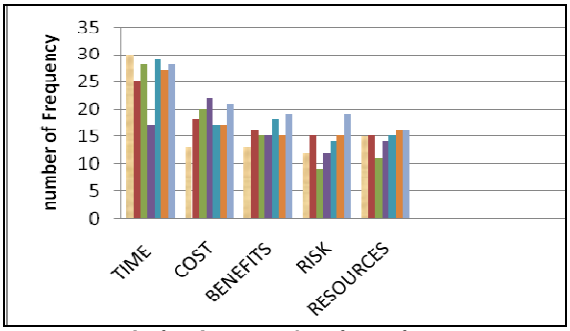


Figure 18. The five decision making factors from 2008-2010.

Based to Figure 18 in looking to the seven columns time obviously is the highest factor. In sum, from all the mentioned and presented here the time is the more important factors, but before we go to the final conclusion it is more better and powerful to present this in a small model since one look equal thousand (a lot of words) this followed in the next section.

For the seven periods from 1990-2010 time and cost factors appeared to be more significant of the DMP, there is a say *"Time is Gold"*. Whereas, for looking for all the DM factors: Time, cost, benefits, risk and resources, were the more important than other factors, which give the decision makers a good idea about inserting and not ignoring those relevant factors in DMP. This will not mean forgetting the other factors, if the decision makers can look for all nine factors it will be better, but if they want to process their decision with the relevant ones only, they can choose what mentioned before and presented in the figures 11,12,13,14,15,16,17 and 18.

VI. PROPOSED MODEL FOR THE DECISION MAKING FACTORS

From all the previous sections a proposed model can be presented for the nine attributes, while this needs other researches to insure it. The model will be in two groups for the factors as independent variables relating to the process of decision making, which is another issue that will help the decision makers in different levels to support them to come with better decisions.

Note: the important group for the five decision making factors linked with normal row, while the second group linked in discrete row in the following Figure 19.

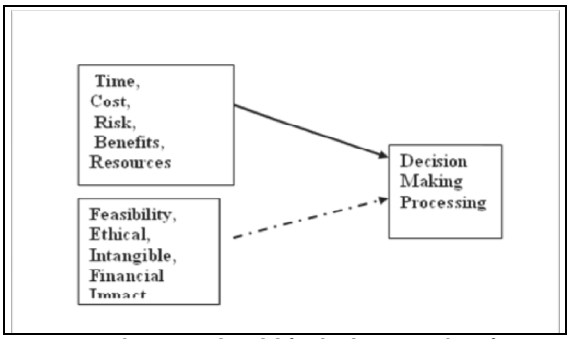


Figure 19. The proposed model for the decision making factors.

VII. CONCLUSION AND FUTURE RESEARCH

Basically researchers help decision makers in decision support systems (DSSs) and had noticed that the decision making processing is the gap in making bad decisions in organizations, for that they presented different ways in processing decisions and referring it to the use of systematic way. Before the processing, this research focus the light on the decision making factors in order to come out with better decisions for multi-decision makers (different level of management and normal users).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Firstly from this qualitative study the factors of decision making are very important in decision making processing, and valuable to the decision makers.

Secondly the factors of decision making can be categorized into two groups: the major (important) group which consists of five factors: cost, time, risk, resources, and benefits, whereby the second group consists of four factors: financial impact, feasibility, intangibles, and ethics.

However the most important factors in is the time, but to rank these factors is not easy here and need other researches which can lead us to end this work with the future researches.

Decision making factors still need more research to be conducted, a comprehensive model verifying all the factors as it help in decision making processing and produce more powerful results, beside using the technology systems as the computer-based information systems (CBIS) in decision making in organizations which will help all humanity to adapt the solution to another areas.

ACKNOWLEDGMENT

The authors wish to acknowledge the reviewers in IJCSIS technical committee for valuable comments, and thank them for selecting this paper for 'Best Paper' Category.

REFERENCES

[1] R. P. Lourence and J. p. Costa, "Incorporating citizens` views in local policy decision making processes," *Decision Support System*, vol.43, pp. 1499-1511, 2007.

[2] K. Haider, J. Tweedale, P. Urlings and L. Jain, "Intelligent decision support system in defense maintenance methodologies," in *International Conference of Emerging Technologies, ICET '06*, pp. 560-567, 13-14 Nov. 2006.

[3] S. M. White, "Requirements for distributed mission-critical decision support systems," *in Proceedings of the 13th Annual IEEE International Symposium & Workshop on Engineering of Computer-Based Systems (ECBS'06)*,Washington, D.C, PP. 123-129, 27-30 March 2006.

[4] R. Luecke, "*Harvard Business Essentials*: *Decision Making 5 Steps to Better Results"*, Boston: Harvard Business School Press. 2006, PP. 47-49.

[5] E. Turban, J. Aronson, T.-P. Liang and R. Sharda, "Decision Support and Business Intelligence Systems." 2007, Prentice Hall, New Jersey,; 8th ed. Prentic-Hall, Newjersy: Pearson: PP. 9-17.

[6] G. E. Vlahos, T. W. Ferratt and G. Knoepfle, "The use of computer-based information systems by German managers to support decision making," in *Information & Managment*, Vol. 41, No. 6, PP. 763-779, 2004.

[7] A. S. Kelton, R. R. Pennington and B. M. Tuttle, "The effects of information presentation format on judgment and decision making: a review of the information systems research," *Journal of Information Systems*, Vol. 24, No. 2, pp. 79-105, 2010.

[8] D. L. Olson and D. D. Wu, "Multiple criteria analysis for evaluation of information system risk," *Asia-Pacific Journal of Operational Research*, Vol. 28, pp. 25-39, 2011.

[9] S. Nowduri, " Management information systems and business decision making: review, analysis, and recommendations," *Journal of Management and Marketing Research*, Vol. 7, PP. 1-8, 2011,

[10] C. Gonzalez, "Decision support for real-time dynamic decision making tasks," *in Organizational Behavior & Human Decision Processes*, Vol. 96, PP. 142–154, 2005a.

[11] S. S. Posavac, F. R. Kardes and J. Josko Brakus, "Focus induced tunnel vision in managerial judgment and decision making: The peril and the antidote," *Organizational Behavior and Human Decision Processes*, Vol. 113, No. 2, PP. 102-111, 2010.

[12] D-J. Kang, J. H. Park and S-S. Yeo , "Intelligent decision-making system with green pervasive computing for renewable energy business in electricity markets on smart grid," *EURASIP Journal on Wireless Communications and Networking*, Hindawi, Vol. 2009, PP. 1-12, 2009.

[13] R. Velik, G. Zucker and D. Dietrich,, "Towards automation 2.0: a neurocognitive model for environment recognition, decision-making, and action execution," *EURASIP Journal on Embedded Systems*, Hindawi, Vol. 2011, PP. 1-11, 2011.

[14] T. L. Lewis, *R. D. Spillmanand and M. Alsawwaf*, "A software engineering approach to the documentation and development of an international decision support system," *Journal of Computing Sciences in Colleges*, Vol. 26, No. 2, PP. 238-245, 2010.

[15] K. Ramprasad and R. Bhattacharya, "State-of-art in regulatory decision making process for a Nuclear Fuel Cycle Facility," *2nd International Conference on Reliability, Safety and Hazard (ICRESH), 2010*, pp.213-218, 14-16 Dec. 2010.

[16] R. S. Wilson and J. L. Arvai, " When less is more: How affect influences preferences when comparing low and high-risk options," *Journal of Risk Research*,Vol. 9, No. 2, PP. 165-178, 2006.

[17] Y-L. Lee and F-H. Huang, "Recommender system architecture for adaptive green marketing," *Expert Systems with Applications: An International Journal*, Vol. 38, No. 8, PP. 9696-9703, 2011.

[18] A. Ferrandez and J. Peral, "The benefits of the interaction between data warehouses and question answering," in *Proceedings of the 2010 EDBT/ICDT Workshops*, vol. 426 of the ACM International Conference Proceeding Series, 2010.

[19] G. Montibeller, L. A. Franco, E. Lord and A. Iglesias, " Structuring resource allocation decisions: A framework for building multi-criteria portfolio models with area-grouped projects," *European Journal of Operational Research*, Vol. 199, No. 3, PP. 846–856, 2009.

[20] T. T. Lee, "Optimizing IT process management," *ACM SIGSOFT Software Engineering Notes*, Vol. 35, No. 4, PP. 1-10, July 2010

[21] T. Klein and A. Thomas, "Opportunities to reconsider decision making processes due to Auto-ID," *Int. J. Production Economics*, Vol. 121, No. 9, PP. 99-111, 2009.

[22] Y- F. Su. And C. Yang, " A structural equation model for analyzing the impact of ERP on SCM," *Expert systems with application*, Vol. 37, PP. 456-469, 2010.

[23] W. Kim, O-R. Jeong, C. Kim and J. So, "The dark side of the Internet: Attacks, costs and responses," *Information Systems*, Vol. 36, No. 3, PP. 675-705, 2011.

[24] A-S. Cheng and K. R. Fleischmann, " Developing a meta-inventory of human values," *Proceedings of the 73rd Annual Meeting of the American Society for Information Science and Technology (ASIS&T)*, 2010, Pittsburgh, PA.

[25] C. L. Jurkiewicz and R. A. Giacalone, "A Values Framework for Measuring the Impact of Workplace Spirituality on Organizational Performance," *Journal of Business Ethics*, Vol. 49, PP. 129–142, 2004.

[26] Y-F. Tseng and T-Z. Lee, "Comparing appropriate decision support of human resource practices on organizational performance with DEA/AHP model," *Expert Systems with Applications*, Vol. 36, No. 3, pp. 6548-6558, 2009.

[27] R. L. Keeney, "Stimulating creative design alternatives using customer values", *IEEE Transactions on Systems Man and Cybernetics Part CApplications and Reviews*, Vol. 34, No. 4, 2004, pp. 450-459.

[28] J. Sarkis and R. P. Sundarraj, "Evaluation of enterprise information technology: A decision model for high-level consideration of strategic and operational issues," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 2, pp. 260–273, Mar. 2006.

[29] L. Kanapeckiene, A. Kaklauskas , E. K. Zavadskas and S. Raslanas, "Method and system for multi-attribute market value assessment in analysis of construction and retrofit projects," *Expert Systems with Applications*, Vol. 38, PP. 14196–14207, 2011.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[30] P. Polet, F. Vanderhaegen, and P. Millot, " Human behavior analysis of barrier deviations using a benefit-cost-deficit model," *Advances in Human-Computer Interaction*, vol. 2009, Article ID 642929, 10 pages, 2009.

[31] U. Sekaran, " Research Method for Business, A Skill Building Approach." 2003, (forth ed). USA: John Wiley & Sons, Inc, PP. 292-296.

Mohammed Suliman Al-Shakkah Received the B.Sc degrees in Maths from yarmouk university in 1998, MSc in Information Technology (IT) from Universiti Sins Malaysia (USM) in 2007, he is vice-dean and lecturer from (2009-2011) in Alghad International Colleges for Health and Medical Sciences University in Kingdom of Saudi Arabia. He is a candidate PhD student in the final stage, started in 2007 Universiti Utara Malaysia (UUM), interested in decision support system (DSS), decision processing for managers in organizations with structuaral equation modeling (SEM) technique, adoption, acceptance and barriers use of computer-based information system (CBIS) in developing countries.

Dr. Wan Rozaini Received the B.Sc degrees in Physics from Universiti Sins Malaysia (USM) in 1982, PG Diploma in Systems Analysis for public Sector from Universiti of Aston in 1983 in UK. She received MSc, ORSA at UK in 1984. PHD, MIS from Universiti of Salge in UK 1996. Now she Associate professor in Universiti Utara Malaysia and Director of ITU- UUM, ASP COE for Rural ICT development.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Role Based Authentication Schemes to Support Multiple Realms for Security Automation

Rajasekhar.B.M & Dr.G.A.Ramachandra

Abstract*—Academy Automation implies to the various different computing hardware and software that can be used to digitally create, manipulate, collect, store, and relay Academy information needed for accomplishing basic Operation like admissions and registration to finance, student and faculty interaction, online library, medical and business development. Raw data storage, electronic transfer, and the management of electronic business information comprise the basic activities of an Academy automation system. The main aim of this work was to design and implement Multiple Realms Authentication where in each realm authentication can be implemented by using Role Based Authentication (RBA) System, where in each user has certain roles allotted to him/her which defines the user's limits and capabilities of making changes, accessing various areas of the software and transferring/allotting these roles recursively. Strict security measures had kept in mind while designing such a system and proper encryption and decryption techniques are used at both ends to prevent any possibility of any third party attacks. Further, various new age authentication techniques like OpenID and WindowsCardSpace are surveyed and discussed to serve as a foundation for future work in this area.

. Index Terms - RBA, Encryption/Decryption, OpenID, WindowsCard-Space.

I. INTRODUCTION

Starting in the 1970s, computer systems featured multiple applications and served multiple users, leading to heightened awareness of data security issues. System administrators and software developers alike focused on different kinds of access control to ensure that only authorized users were given access to certain data or resources. One kind of access control that emerged is role-based access control (RBAC). A role is chiefly a semantic construct forming the basis of access control policy. With RBAC, system administrators create roles accordingly to the job functions performed in a company or organization, grant permissions (access authorization) to those roles, and then assign users to the roles on the basis of their specific job responsibilities and qualifications "Role-based access control terms and concepts". A role can represent specific task competency, such as that of a physician or a pharmacist. A role can embody the authority and responsibility of, say , a project supervisor. Authority and responsibility are distinct from competency. A person may be competent to manage several departments but have the responsibility for only the department

actually managed. Roles can also reflect specific duty assignments rotated through multiple users for example, a duty physician or a shift manager. RBAC models and implementations should conveniently accommodate all these manifestations of the role concept. Roles define both the specific individuals allowed to access resources and the extent to which resources are accessed. For example, an operator role might access all computer resources but not change access permissions; a security-officer role might change permissions but have no access to resources; and an auditor role might access only audit trails. Roles are used for system administration in such network operating systems as Novell's NetWare and Microsoft's Windows NT.

In this article present a comprehensive approach to RBAC on the Web. We identify the user-pull and server-pull architectures and analyze their utility. To support these architectures on the Web, for relatively mature technologies and extend them for secure RBAC on the Web. In order to do so, to make use of standard technologies in use on the Web: cookies [Kristol and Montulli 1999; Moore and Freed 1999], X.509 [ITU-TRecommendation X.509 1993; 1997; Housley et a1. 1998], SSL (Secure Socket Layer [Wagner and Schneier 1996; Dierks and Allen 1999]), and LDAP (Lightweight Directory Access Protocol [Howes et a1. 1999]), and LDAP (Lightweight Directory Access Protocol (LDAP) directory service already available for the purpose of web mail authentication of Sri Krishna Devaray University, Anantapur users has been used to do the basic Authentication. The client can request the application server for any web application which will ask for the user credentials which will be verified in the LDAP server through an J2EE[17] Module. On successful verification, the authorization module will contact the user role database and fetch the roles for that user. In case of return of multiple roles user will be given the authorization of all the roles. The access to the application will be on the basis of privilege of the role of that particular user. The role database is implementing in Oracle databse. On successful authentication, the Authentication and authorization module which has been developed for this purpose is called and the role for the user is retrieved. Privileges are granted to roles and interns are granted to users.

The overall database server and application server is considered for possible attacks. The proposed schema is given figure 2. The database server and authentication server are in a private network and separated from the user network by a firewall. These servers can be accessed only through application server, i.e through the authentication and authorization module. Application server has an interface in the private network but can avail only the specific service which has been explicitly allowed in the firewall. Application server has another interface which is part of user network with a firewall to restrict the clients only to the desired service.

The information flow security has been taken care by secure http. The J2EE Application server has the support for HTTPS which was configured to make sure that data passing to

1. Dept of Computer Science & Technology, S.K. University, Anantapur
rajasekhar3@gmail.com, chandragar@yahoo.com

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and from Application server is encrypted. From the Application Server, a digital certificate in SSL [23] (Secure Socket Layer) has been generated. This needs to be installed on the client machine for server identity verification. Similarly client certificate can also be generated from the J2EE which can be used in the client which will update sensitive data. Such operation will be denied without client certificate.

II. LITERATURE REVIEW

A large number of research papers are published in the area of Role Based Authentication In [5] Raymond emphasized the purpose of Role Based Authentication. Authorization architecture for authorizing access to resource objects in an object-oriented programming environment is discussed in this paper. In one distributed environment, the permission model of JAAS (Java Authentication and Authorization Service) is replaced or enhanced with role-based access control. Thus, users and other subjects (e.g., pieces of code) are assigned membership in one or more roles, and appropriate permissions or privileges to access objects are granted to those roles. Permissions may also be granted directly to users. Roles may be designed to group users having similar functions, duties or similar requirements for accessing the resources. Roles may be arranged hierarchically, so that users explicitly assigned to one role may indirectly be assigned to one or more other roles (i.e.,descendants of the first role). A realm or domain may be defined as a namespace, in which one or more role hierarchies are established.

Robert et al in [6] discussed about Methods, systems, and computer program products are disclosed for protecting the security of resources in distributed computing environments.The disclosed techniques improve administration and enforcement of security policies. Allowed actions on resources, also called permissions, (such as invocations of particular methods, read or write access of a particular row or perhaps a particular column in a database table, and so forth) are grouped, and each group of permissions is associated with a role name. A particular action on a particular resource may be specified in more than one group, and therefore may be associated with more than one role. Each role is administered as a security object. Users and/or user groups may be associated with one or more roles. At run-time, access to a resource is protected by determining whether the invoking user has been associated with (granted) at least one of the roles required for this type of access on this resource.

In [7] Dixit et al discussed about an actor is associated with a role, a policy type is associated with the role, and a role scope is associated with the role. One or more values are received for one or more corresponding context parameters associated with the actor. A request for access to a resource is received from the actor. A policy instance is determined based on the policy type and the one or more values for the one or more corresponding context parameters associated with the actor. One or more actor-role scope values are determined based on the role scope and the one or more values for the one or more corresponding context parameters associated with the actor. A response to the request is determined based on the policy instance and the actor-role scope values.

Bindiganavale and Ouyang, in [8] presents the most challenging problems in managing large web-applications is the complexity of security administration and user-profile management. Role Based Access Control (RBAC) has become the predominant model for advanced access control because it reduces the complexity and cost of administration. Under RBAC, security administration is greatly simplified by using roles, hierarchies and privileges, and user management is uncomplicated by using LDAP API specification within the J2EE application. System administrators create roles according to the job functions performed in an organization, grant permissions to those roles, and then assign users to the roles on the basis of their specific job Responsibilities and qualifications.

A wireless networks proliferate, web browsers operate in an increasingly hostile network environment. The HTTPS protocol has the potential to protect web users from network attackers, but real-world deployments must cope with misconfigured servers, causing imperfect web sites and users to compromise browsing sessions inadvertently. Force HTTPS is a simple browser security mechanism that web sites or users can use to opt in to stricter error processing, improving the security of HTTPS by preventing network attacks that leverage the browser's lax error processing. By augmenting the browser with a database of custom URL rewrite rules, Force HTTPS allows sophisticated users to transparently retrofit security onto some insecure sites that support HTTPS. We provide a prototype implementation of Force HTTPS as a Firefox browser extension [9].

A comparison of a simple RBAC model and a group Access Control List(ACL) mechanism by Barkley [10] shows that even the simplest RBAC model is as effective in its ability to express access control policy. An RBAC system with special features (which are not possible with ACLs) will be even more effective.

III. OBSERVATIONS AND PROBLEM DESCRIPTION

The whole Collage Academy automation consists of many sections viz. Student Affairs, Academic Section, Research and Development, Training and Placement, Finance and Accounts etc. For example if IPS Academy wants to integrate with different Academy's like Indore Institute of Science & Technology then in that case we can implement Multiple Realm Authentication System. Different individuals in IPS Academy, Indore should be given access to different aspects of the systems based on their clearance level. For e.g. the Assistant Registrar of Student Affairs should have full access to all the options of Student Affairs database but not that of the Academic Section database. However, provisions have to be made so that he/she is able to perform some student affairs related queries to the student affairs database. Similarly, a student must have read-only access to his/her information in the official records and modifying capabilities some of his/her details in the training and placement section database. This calls for a role-based approach to access the databases. Each person has a certain role attached to it. This role corresponds to the areas of the work his login account can access. If a violation occurs, the user is immediately logged out.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In this work the design and implementation of the Role Based Authentication Schemes to Support Multiple Realms for Security Automation is described, developed at the IPS Academy, Indore as an Java, J2EE [2005] web application in JSP server side code, HTML, and JavaScript for use on the Internet. The purpose work to deploy a cost-effective, web-based system that significantly extends the capabilities, flexibility, benefits, and confidentiality of paper-based rating methods while incorporating the ease of use of existing online surveys and polling programs.

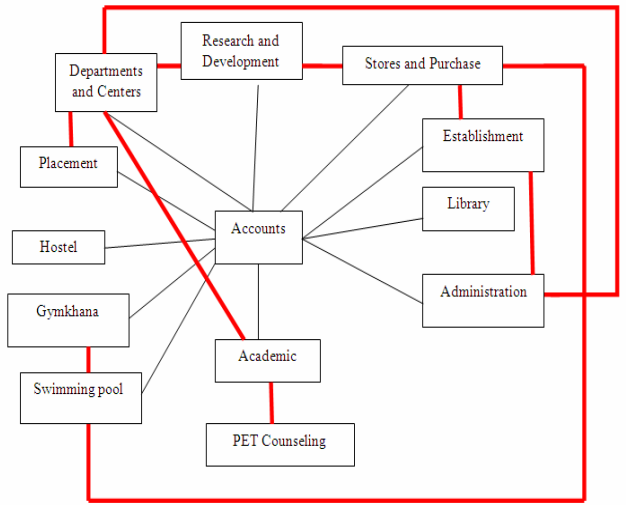


Figure 1: Basic Architecture of Academy

A. Problem Issues And Challenges

The following Problems are as Follows:-

1) The information line must be completely secured.

2) Proper Encryption must be used for storing the Password for the User.

3) The authorization token which is stored on the client side has to be encrypted so that the client cannot modify his authorization clearance level.

4) Each userid-role mapping should have an expiry date beyond which it will be invalid.

5) Role Scoping: Local and Global Roles

6) In each role, we have to have an owner. Normally the role will map to the user id of the owner. The owner can change the mapping and can specify the time period of this change. The newly mapped user is not the owner and so cannot change the ownership, but maybe allowed to map again. For example, HODCSE is the role and the owner's user id is" Ram". Normally, HODCSE maps to Ram. When Prof. Ram goes on leave, he fills up some form electronically and this triggers (among other things) a role change of HODCSE to the user he designates, say Prof.Shayam. Now" Ram" is going on leave till 4/7/2010, so the changed mapping is till 4/7/2010 (to"pshayam"; specified by" Ram" in the form he filled up). Now due to an emergency, "pshayam" had to leave station on 4/7/2010, making Prof manoj the Head. Since" pshayam" is not the owner, he cannot change the validity date beyond 4/7/2010 and"Ashish" takes over the HODCSE role till 4/7/2010. On 5/7/2010 (or the next query of the role), the role remaps to"Ram". Other cases (like"Ram" having to overstay beyond 4/7) can be handled by the administrator.used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

7) We need to write N no.of authenticators based on requirements.

8) Based on role name(which we can get it from login page), we can create associate authenticators through reflection api for authenticating username and password.

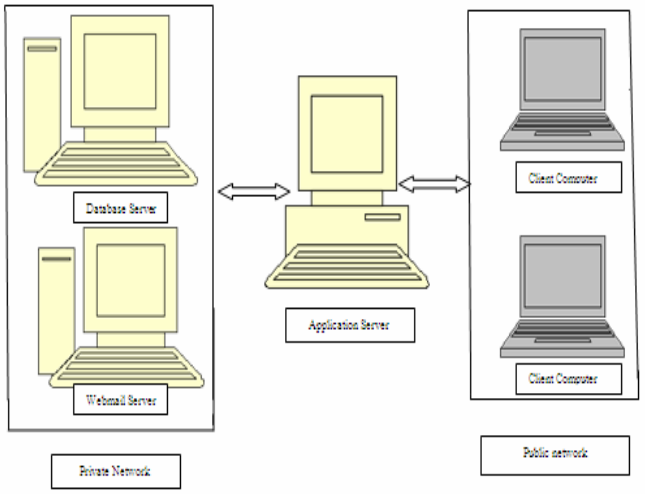


Figure 2: System and Server Security

IV. METHODOLOGIES

1) We have 2 sets of Roles:

Global Roles: These refer to the roles which are common to the entire applications viz. root, Director. Their Role IDs are of single digit: 0, 1, and 2 etc.

Local Roles: These are roles which are specific to a module. For E.g. for Student Affairs, the roles of Assistant Registrar, Academy in charge. Their IDs are of the Form: 10, 11, 12 ... 110 etc. where first digit identifies the application to which all of them are common.

2) There is a Global role to role id mapping table.

3) Also there is a local mapping table for each section. Insertion/modification or deletion of any entry in the local table generates a Microsoft SQL trigger for its 'encoded' entry addition in the global table.

Below table describes about the Realm association to the domain as such, each domain is associated to unique domain. And where administrator can have to privileges to active or Inactive domain level.

For Example: Realm→Domain-→Users

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Realm ID	Realm Name	Active/Inactive
1	Academy Realm	A
2	XXX Realm	A
3	YYY Realm	A

TABLE 1: REALMS

Below is the table of which unique role id has been assigned to specific role. So as an administrator can have Full privilege to all domains and the rest has to login with their role id's.

Role	Role ID
Administrator	0
Student	1
Assistant Registrar(Student Affairs)	10
Assistant Registrar(Academic)	20
Assistant Registrar(R&D)	30
Assistant Registrar(Finance)	40
Registrar	3
Director	4
Head of Depts	5

TABLE 2: VARIOUS ROLES AND THEIR IDs

Below Table Describes about users association to Realm with and unique Realm ID. And whereas same user id is uniquely associated to user name. whereas mapping goes in such a way like.

Example: User Name→User Id→ Realm ID

User_name	User_id	Realm ID
root	11	1
rajasekhar	22	2
test	33	3
admin	55	3
michael	66	2
tang	88	2

TABLE 3: USER NAME ID RELATION

Below Table Describes about users association to Realm with and unique Realm ID.And whereas same user id is uniquely associated to user name. whereas mapping goes in such a way like.

In this case each and every users have validate dates of which user can access the domain in the associated realm. If so the users cross there validity dates he nowhere access the associated realm /System.

Example: User Name→User Id→ Valid Up to→Realm ID

S_no	User_id	Role_id	Valid_from	Valid_upto
1	11	6	2008-01-01	2011-12-01
2	11	5	2008-03-01	2011-03-01
3	22	1	2003-07-02	2005-07-10
4	33	4	2008-08-04	2011-09-15
5	66	3	2009-10-10	2011-12-12
6	88	20	2010-08-08	2012-08-08

TABLE 4: USER ROLE RELATION

A web interface which is accessed by any member and is used to assign his role to any other member for a specified period. The role validity period of the other person cannot exceed the validity period of the assigner. So, whenever a role has to be transferred, an entry is made in the user role relation table corresponding to the user ID of the assigned person and it is made sure that the validity period of the assigned is less than the validity period of assigner from the same user role relation table

A. Database Table Structure

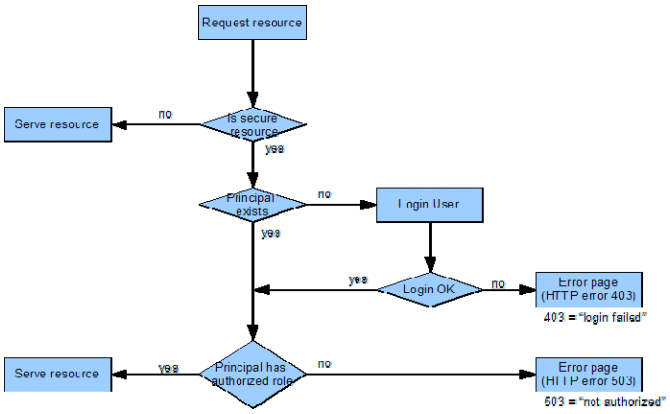
We will have a common login page for all the sections of the Academy Automation. The looks up table of the corresponding IDs are shown in table 1, 2 , 3 & 4.

B. Java, J2EE Authentication

Now, each webpage has a small Jsp & Servlet and Java code which expects to read the system cookie of a specified number of roles before displaying the page. If unsuccessful, this page re-directs the user to the logout page and deletes the session cookies else the corresponding web page is displayed.

So what happens when you access a secured web application resource? The diagram below shows the typical rundown of accessing a web resource with security enabled.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



And now in verbose mode: the usual path is 1) check if the resource is secured; 2) check if the requesting user has been authenticated; 3) check if the authenticated user is properly authorized to access the requested resource and 4) serve the requested resource. If the user has not been authenticated yet, walk through the Login dialog. If anything is out of order, display the corresponding error page. Or, if the resource is not secure, skip all previously mentioned steps and serve the resource right away.

We must create a Forms authentication login system the supports roles. The process of creating the authentication ticket and the cookie has to be stored under the right name – the name matching the configured name for Forms authentication root config file. If these names don't match, servlet wouldn't find the authentication ticket for the Web application and force a redirect to the login page. The authentication module which is imported at the beginning of every jsp page. In the login page we can display username and password along with domain names which we get it from REALMS table. While login into site, end user has to select any one of the domain in login page.

The below method can convert password into hash. Here I used one-way hash algorithm and that makes a unique array of characters.

FormsAuthentication.HashPasswordForStoringInConfigFile(Password);

We do one other thing with our passwords: we hash them. Hashing is a one-way algorithm that makes a unique array of characters. Even changing one letter from upper-case to lower-case in your password would generate a completely different hash. We'll store the passwords in the database as hashes, too, since this is safer. In a production environment, we'd also want to consider having a question and response challenge that a user could use to reset the password. Since a hash is one-way, we won't be able to retrieve the password. If a site is able to give our old password to us, I'd consider steering clear of them unless you were prompted for a client SSL certificate along the way for encrypting your pass phrase and decrypting it for later use, though it should still be hashed.

C. *Securing Directories with Role-Based Forms Authentication*

In order to make the role Based authentication work for Forms Authentication, it is required to have a configuration file in web application root. In authentication setup, this particular config file must be in Web Application's document root.

MyFilterSecurity contains the definitions of the secured resources. Let's take a look at the XML configuration first:

```
<property name="objectDefinitionSource">
<value>CONVERT_URL_TO_LOWERCASE_BEFORE_
COMPARISONPATTERN_TYPE_APACHE_ANT
/secure/admin/*=ROLE_ADMIN
/secure/app/*=ROLE_USER
</value>
</property>
```

In the above configuration, "secured resources" are called "object definitions" (it is a rather generic sounding name because our research can be also used to control access to method invocations and object creations, not just web applications). The thing to remember here is that "objectDefinitionSource" should contain some directives and the URL patterns to be secured, along with the roles who have access to those URL patterns.

D. *Conditionally Showing Controls With Role-Based Forms Authentication*

The IPrincipal interface, which the GenericPrincipal class we used above implements, has a method called "IsInRole()", which takes a string designating the role to check for. So, if we can only want to display content if the currently logged-on user is in the "Administrator" role.

```
<html>
<head>
<title>Welcome</title>
<script language="javascript">
Function isUserRole()
{
if (User.IsInRole("Administrator"))
AdminLink.Visible = true;
}
</script>
</head>
<body>
<h2>Welcome</h2>
<p>Welcome, anonymous user, to our web site.</p>
<p><a href="/AdminLink "> Administrators </a>
</body>
</html>
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

E. Configuring Multiple Realms

In order to support the multiple realms to existing approach then we can write sql insert query script for inserting N no.of realms or domains for adding into REALMS table.

V. COMPARISION OF EXISTING AND CURRENT APPROACH

The main aim of Role Based Authentication Schemas for Security Automation Publication[24], work was to design and implements a Role Based Authentication (RBA) System wherein each user has certain roles allotted to him/her which defines the user's limits and capabilities of making changes, accessing various areas of the software and transferring/allotting these roles recursively. In the existing publication [24], it will apply only for one realm authentication. For example consider two domains, D1 and D2. Where in D1 domain consists of the whole College students and staff and D2 domain consists of only Distance College students and staff. In the existing publication approach can authenticate either D1 users or D2 users but it can't authenticate both the domain users.

To overcome the existing problem, and introduced multiple realms authentication approach. In which we can authenticate more than one domain user. We can categorize it into two realms, R1 and R2. We can store D1 users info into realm R1 and D2 users info into realm R2. We can categorize it into N (R1,R2,R3….Rn) no of realms. See more details in METHODOLOGIES section.

VI. CONCLUSION

The research problem and goal of the Academy Automation is to design a highly secure and efficient framework based on SOA keeping all policies on note for minimum data redundancy and providing an option for authentication of different realms with efficient security, the work revolved around designing a plug in for secure role based authentication. Presently the authentication is based on the traditional user id and password based approach and can be authenticated against multiple realms, but it is suggested in the report, future work can be done to incorporate various new-age techniques such as OpenID…etc.

REFERENCES

[1] William Stallings, "Cryptography and Network Security Principles and Practices", 3rdEdition,Prentice Hall, 2003.

[2] Eric Cole, Ronald L. Krutz, James Conley, "Network Security Bible", 2nd Edition, Wiley Publication, 2005.

[3] Yih-Cheng Lee, Chi-Ming Ma and Shih-Chien Chou," A Service-Oriented Architecture for Design and Development of Middleware," Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC05) 0-7695- 2465-6/05

[4] Wagner, David; Schneier and Bruce,"Analysis of the SSL 3.0 Protocol," The Second USENIX Workshop on Electronic Commerce Proceedings, USENIX Press. Nov 1996.

[5] Ng, Raymond K."Distributed capability-based authorization architecture using roles"2004.

[6] Robert Jr., Howard High (Round Rock, TX, US), Nadalin, Anthony Joseph (Austin, TX, US), Nagaratnam,Nataraj (Morrisvile, CA, US) ,"Role-permission model for security policy administration and enforcement" 2003.

[7] Dixit, Royyuru (Wilmington, MA, US), Hafeman, Joseph Edward (Holliston, MA, US), Vetrano, Paul Michael(Franklin, MA, US), Spellman, Timothy Prentiss (Framingham, MA, US), "Role-based access in a multi customer computing environment",2006.

[8] Vinith Bindiganavale and Jinsong Ouyang, Member, IEEE [2001]." Role Based Access Control in Enterprise Application – Security Administration and User Management"

[9] Collin Jackson and Adam Barth," Force HTTPS: Protecting High Security Web Sites from Network Attacks"

[10] Barkley J., "Comparing Simple Role Based Access Control Models and Access Control Lists", Second ACM workshop on Role-based Access Controls, 1997.

[11] Sanin, Aleksey (Sunnyvale, CA, US),"Web service security filter"

[12] Chung, Hyen V. (Round Rock, TX, US), Nakamura, Yuhichi (Yokohama-Shi, JP), Satoh, Fumiko (Tokyo, JP), "Security Policy Validation for Web Services".

[13] Kou, Wei Dong (Pokfulam, HK) ,Mirlas, Lev (Thornhill, CA), Zhao and Yan Chun (Toronto, CA)," On Secure session management and authentication for web sites",2005.

[14] Akram Alkouz and Samir A. El-Seoud (PSUT) Jordan," Web Services Based Authentication System For E-Learning", 2005.

[15] Thomas Price, Jeromie Walters and Yingcai Xiao, "Role –Based Online Evaluation System", 2007.[15] Srinath Akula, Veerabhadram Devisetty , St. Cloud, MN 56301." Image Based Registration and Authentication System", 2002.Rivest, Shamir and Adelman,"RSA public-key encryption".

[16] Microsoft,"Asp.net".Website:http://www.asp.net/index.html.

[17] Netfilter Core team, Iptables, an userspace packet filtering program

[18] Website:http://www.netfilter.org/projects/iptables/index.html.

[19] Digital Certificates for Internet Security and Acceleration Server 2004, for Microsoft Forefront Threat Medium Business Edition, or for Windows Essential Business Server 2008. Website: http://support.microsoft.com/kb/888716

[20] OpenID Foundation. Website: openid.net/

[21]Microsoft,"VisualStudioJ2EE".
Website:http://msdn.microsoft.com/vstudio/

[22] Microsoft Corporation. Website: www.passport.net/

[23] OpenSSL team. Website: http://www.openssl.org/

[24] Role Based Authentication Schemes for Security Automation.

AUTHOR BIOGRAPHIES

Rajasekhar.B.M holds a M.Sc in Computer Science from S.K.University, AP-India. He is currently pursuing M.Phil in Computer Science, S.K. University, AP-India. And he is also currently Associate-Projects in Cognizant Technologies India Pvt Ltd. His research interest includes network security, web security, routing algorithms, client-server computing and IT based education.

Dr.G.A.Ramachandra obtained his Ph.D in Mathematics from S.K.University, AP-India. He is currently Associate Professor in the Dept of Computer Science and Technology, S.K.University, AP-India. His area of interest is on Computer Networks, Network Security and Image Processing. In His tenure of Headshiphe was co-ordinator for establishing On line Counseling Center at Sri Krishnadevaraya Universityas part of Andhra Pradesh State Council of Higher Education. He Published 10 Papers for National /International Journals. He attended for 2-National Conferences.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Parallel Edge Projection and Pruning (PEPP) Based Sequence Graph Protrude Approach for Closed Itemset Mining

kalli Srinivasa Nageswara Prasad
Research Scholar in Computer Science
Sri Venkateswara University, Tirupati
Andhra Pradesh , India.
.

Prof. S. Ramakrishna
Department of Computer Science
Sri Venkateswara University, Tirupati
Andhra Pradesh , India.
.

Abstract: **Past observations have shown that a frequent item set mining algorithm are supposed to mine the closed ones as the end gives a compact and a complete progress set and better efficiency. Anyhow, the latest closed item set mining algorithms works with candidate maintenance combined with test paradigm which is expensive in runtime as well as space usage when support threshold is less or the item sets gets long. Here, we show, PEPP, which is a capable algorithm used for mining closed sequences without candidate. It implements a novel sequence closure checking format that based on Sequence Graph protruding by an approach labeled "Parallel Edge projection and pruning" in short can refer as PEPP. A complete observation having sparse and dense real-life data sets proved that PEPP performs greater compared to older algorithms as it takes low memory and is more faster than any algorithms those cited in literature frequently.**

Key words – Data Mining; Graph Based Mining; Frequent itemset; Closed itemset; Pattern Mining; candidate; Itemset Mining; Sequential Itemset Mining.

I. INTRODUCTION

Sequential item set mining, is an important task, having many applications with market, customer and web log analysis, item set discovery in protein sequences. Capable mining techniques are being observed extensively, including the general sequential item set mining [1, 2, 3, 4, 5, 6], constraint-based sequential item set mining [7, 8, 9], frequent episode mining [10], cyclic association rule mining [11], temporal relation mining [12], partial periodic pattern mining [13], and long sequential item set mining [14]. Recently it's quite convincing that for mining frequent item sets, one should mine all the closed ones as the end leads to compact and complete result set having high efficiency [15, 16, 17, 18], unlike mining frequent item sets, there are less methods for mining closed sequential item sets. This is because of intensity of the problem and CloSpan is the only variety of algorithm [17], similar to the frequent closed item set mining algorithms, it follows a candidate maintenance-and-test paradigm, as it maintains a set of readily mined closed sequence candidates used to prune search space and verify whether a recently found frequent sequence is to be closed or not. Unluckily, a closed item set mining algorithm under this paradigm has bad scalability in the number of frequent closed item sets as many frequent closed item sets (or just candidates) consume memory and leading to high search space for the closure checking of recent item sets, which happens when the support threshold is less or the item sets gets long.

Finding a way to mine frequent closed sequences without the help of candidate maintenance seems to be difficult. Here, we show a solution leading to an algorithm, PEPP, which can mine efficiently all the sets of frequent closed sequences through a sequence graph protruding approach. In PEPP, we need not eye down on any historical frequent closed sequence for a new pattern's closure checking, leading to the proposal of Sequence graph edge pruning technique and other kinds of optimization techniques.

The observations display the performance of the PEPP to find closed frequent itemsets using Sequence Graph. The comparative study claims some interesting performance improvements over BIDE and other frequently cited algorithms.

In section II, most frequently cited work and their limits explained. In section III, the Dataset adoption and formulation explained. In section IV, introduction to PEPP and its utilization

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

for Sequence Graph protruding explained. In section V, the algorithms used in PEPP described. In section V1, results gained from a comparative study briefed and followed by conclusion of the study.

II. RELATED WORK

The sequential item set mining problem was initiated by Agrawal and Srikant , and the same developed a filtered algorithm, GSP [2], basing on the Apriori property [19]. Since then, lots of sequential item set mining algorithms are being developed for efficiency. Some are, SPADE [4], PrefixSpan [5], and SPAM [6]. SPADE is on principle of vertical id-list format and it uses a lattice-theoretic method to decompose the search space into many tiny spaces, on the other hand PrefixSpan implements a horizontal format dataset representation and mines the sequential item sets with the pattern-growth paradigm: grow a prefix item set to attain longer sequential item sets on building and scanning its database. The SPADE and the PrefixSPan highly perform GSP. SPAM is a recent algorithm used for mining lengthy sequential item sets and implements a vertical bitmap representation. Its observations reveal, SPAM is better efficient in mining long item sets compared to SPADE and PrefixSpan but, it still takes more space than SPADE and PrefixSpan. Since the frequent closed item set mining [15], many capable frequent closed item set mining algorithms are introduced, like A-Close [15], CLOSET [20], CHARM [16], and CLOSET+ [18]. Many such algorithms are to maintain the ready mined frequent closed item sets to attain item set closure checking. To decrease the memory usage and search space for item set closure checking, two algorithms, TFP [21] and CLOSET+2, implement a compact 2-level hash indexed result-tree structure to keep the readily mined frequent closed item set candidates. Some pruning methods and item set closure verifying methods, initiated the can be extended for optimizing the mining of closed sequential item sets also. CloSpan is a new algorithm used for mining frequent closed sequences [17]. It goes by the *candidate maintenance-and-test* method: initially create a set of closed sequence candidates stored in a hash indexed result-tree structure and do post-pruning on it. It requires some pruning techniques such as *Common Prefix* and *Backward Sub-Item set pruning* to prune the search space as CloSpan requires maintaining the set of closed sequence candidates, it consumes much memory leading to heavy search space for item set closure checking when there are more frequent closed sequences. Because of which, it does not scale well the number of frequent closed sequences. BIDE **[**26**]** is another closed pattern mining algorithm and ranked high in performance when compared to other algorithms discussed. Bide projects the sequences after projection it prunes the patterns that are subsets of current patterns if and only if subset and superset contains same support required. But this model is opting to projection and pruning in sequential manner. This sequential approach sometimes turns to expensive when sequence length is considerably high. In our earlier literature[27] we discussed some other interesting works published in recent literature.

Here, we bring Sequence Graph protruding that based on edge projection and pruning, an asymmetric parallel algorithm for finding the set of frequent closed sequences. The giving of this paper is: (A) an improved sequence graph based idea is generated for mining closed sequences without candidate maintenance, termed as Parallel Edge Projection and pruning (PEPP) based Sequence Graph Protruding for closed itemset mining. The Edge Projection is a forward approach grows till edge with required support is possible during that time the edges will be pruned. During this pruning process vertices of the edge that differs in support with next edge projected will be considered as closed itemset, also the sequence of vertices that connected by edges with similar support and no projection possible also be considered as closed itemset (B) in the Edge Projection and pruning based Sequence Graph Protruding for closed itemset mining, we create a algorithms for Forward edge projection and back edge pruning(C) the performance clearly signifies that proposed model has a very high capacity: it can be faster than an order of magnitude of CloSpan but uses order(s) of magnitude less memory in several cases. It has a good scalability to the database size. When compared to BIDE the model is proven as equivalent and efficient in an incremental way that proportional to increment in pattern length and data density.

III. DATASET ADOPTION AND FORMULATION

Item Sets I: A set of diverse elements by which the sequences generate.

$$I = \bigcup_{k=1}^{n} i_k$$

Note: ‘I’ is set of diverse elements

Sequence set ‘S’: A set of sequences, where each sequence contains elements each element ‘e’ belongs to ‘I’ and true for a function p(e). Sequence set can formulate as

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$s = \bigcup_{i=1}^{m} < e_i \mid (p(e_i), e_i \in I) >$$

Represents a sequence's' of items those belongs to set of distinct items 'I'.

'm': total ordered items.

$P(e_i)$: a transaction, where e_i usage is true for that transaction.

$$S = \bigcup_{j=1}^{t} s_j$$

S: represents set of sequences

't': represents total number of sequences and its value is volatile

s_j: is a sequence that belongs to S

Subsequence: a sequence s_p of sequence set 'S' is considered as subsequence of another sequence s_q of Sequence Set 'S' if all items in sequence S_p is belongs to s_q as an ordered list. This can be formulated as

If $(\bigcup_{i=1}^{n} s_{pi} \in s_q) \Rightarrow (s_p \subseteq s_q)$

Then $\bigcup_{i=1}^{n} s_{pi} <: \bigcup_{j=1}^{m} s_{qj}$ where $s_p \in S \text{ and } s_q \in S$

Total Support 'ts' : occurrence count of a sequence as an ordered list in all sequences in sequence set 'S' can adopt as total support 'ts' of that sequence. Total support 'ts' of a sequence can determine by following formulation.

$$f_{ts}(s_t) = \mid s_t <: s_p \ (\textit{for each } p = 1..\mid DB_S \mid) \mid$$

DB_S Is set of sequences

$f_{ts}(s_t)$: Represents the total support 'ts' of sequence s_t is the number of super sequences of s_t

Qualified support 'q_s': The resultant coefficient of total support divides by size of sequence database adopt as qualified support 'qs'. Qualified support can be found by using following formulation.

$$f_{qs}(s_t) = \frac{f_{ts}(s_t)}{\mid DB_S \mid}$$

Sub-sequence and Super-sequence: A sequence is sub sequence for its next projected sequence if both sequences having same total support.

Super-sequence: A sequence is a super sequence for a sequence from which that projected, if both having same total support.

Sub-sequence and super-sequence can be formulated as

If $f_{ts}(s_t) \geq$ rs where 'rs' is required support threshold given by user

And $s_t <: s_p \ \textit{for any p value}$ where $f_{ts}(s_t) \cong f_{ts}(s_p)$

IV. PARALLEL EDGE PROJECTION AND PRUNING BASED SEQUENCE GRAPH PROTRUDE

Preprocess:

As a first stage of the proposal we perform dataset preprocessing and itemsets Database initialization. We find itemsets with single element, in parallel prunes itemsets with single element those contains total support less than required support.

Forward Edge Projection:

In this phase, we select all itemsets from given itemset database as input in parallel. Then we start projecting edges from each selected itemset to all possible elements. The first iteration includes the pruning process in parallel, from second iteration onwards this pruning is not required, which we claimed as an efficient process compared to other similar techniques like BIDE. In first iteration, we project an itemset s_p that spawned from selected itemset s_i from DB_S and an element e_i considered from 'I'. If the $f_{ts}(s_p)$ is greater or equal to rs,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

then an edge will be defined between s_i and e_i . If $f_{ts}(s_i) \cong f_{ts}(s_p)$ then we prune s_i from DB_S . This pruning process required and limited to first iteration only.

From second iteration onwards project the itemset S_p that spawned from $S_{p'}$ to each element e_i of 'I'. An edge can be defined between $S_{p'}$ and e_i if $f_{ts}(s_p)$ is greater or equal to rs. In this description $S_{p'}$ is a projected itemset in previous iteration and eligible as a sequence. Then apply the fallowing validation to find closed sequence.

If any of $f_{ts}(s_{p'}) \cong f_{ts}(s_p)$ that edge will be pruned and all disjoint graphs except s_p will be considered as closed sequence and moves it into DB_S and remove all disjoint graphs from memory.

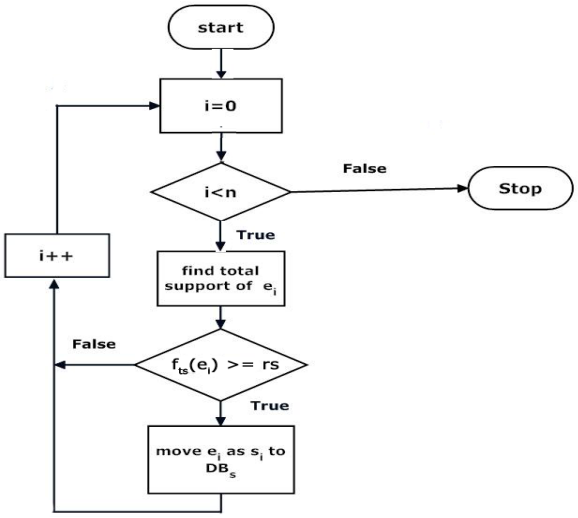
If $f_{ts}(s_{p'}) \cong f_{ts}(s_p)$ and there after no projection spawned then s_p will be considered as closed sequence and moves it into DB_S and remove $s_{p'}$ *and* s_p from memory.

The above process continues till the elements available in memory those are connected through direct or transitive edges and projecting itemsets i.e., till graph become empty.

V. ALGORITHMS USED IN PEPP

This section describes algorithms for initializing sequence database with single elements sequences, spawning itemset projections and pruning edges from Sequence Graph SG.

Figure 1: Generate initial DB_S with single element itemsets



Algorithm 1: Generate initial DB_S with single element itemsets

Input: Set of Elements 'I'.

Begin:

L1: For each element e_i of 'I'

Begin:

Find $f_{ts}(e_i)$

If $f_{ts}(e_i) \geq rs$ then

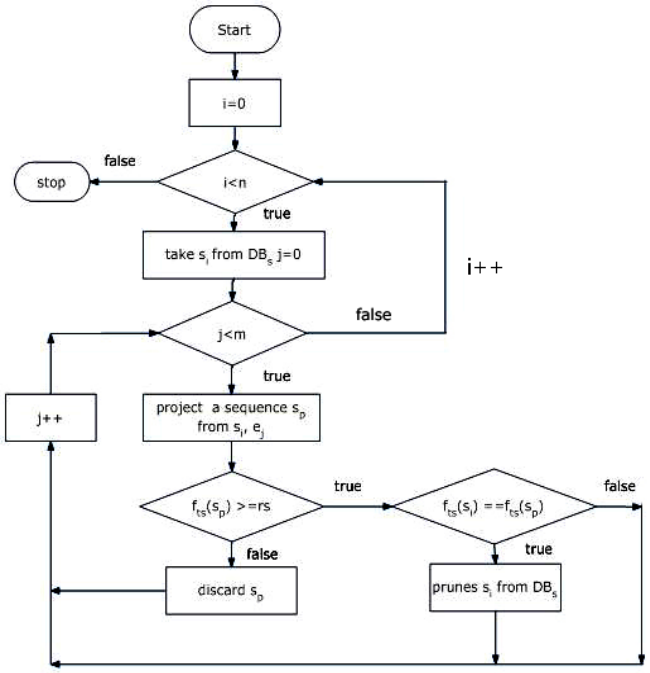
Move e_i as sequence with single element to DB_S

End: L1.

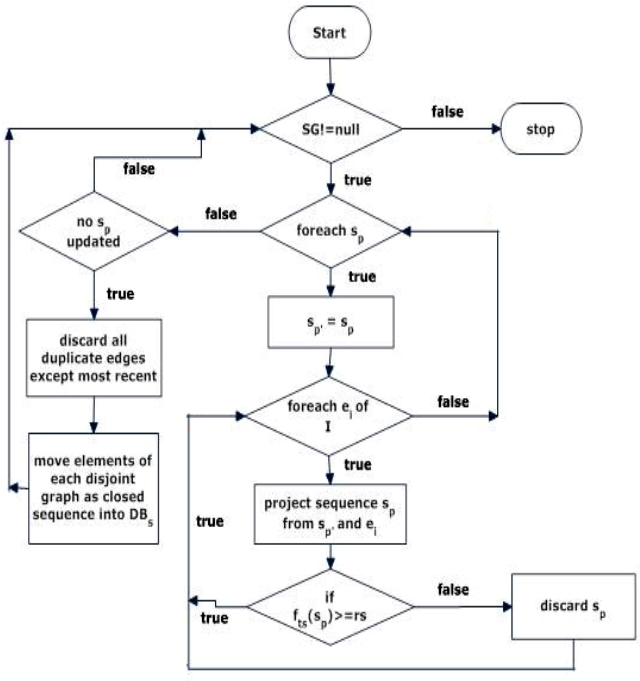
End.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure2: spawning projected Itemsets and protruding sequence graph



(a) First iteration



(b) Rest of all Iterations

Algorithm 2: spawning projected Itemsets and protruding sequence graph

Input: DB_S and 'I';

L1: For each sequence s_i in DB_S
Begin:

L2: For each element e_i of 'I'
Begin:

C1: if $\text{edgeWeight}(s_i, e_i) \geq rs$

Begin:

Create projected itemset s_p from (s_i, e_i)

If $f_{ts}(s_i) \cong f_{ts}(s_p)$ then prune s_i from DB_S

End: C1.
End: L2.
End: L1.

L3: For each projected Itemset s_p in memory

Begin:

$s_{p'} = s_p$

L4: For each e_i of 'I'
Begin:

Project s_p from $(s_{p'}, e_i)$

C2: If $f_{ts}(s_p) \geq rs$

Begin

Spawn SG by adding edge between $s_{p'}$ *and* e_i

End: C2
End: L4

C3: If $s_{p'}$ not spawned and no new projections added for $s_{p'}$

Begin:

Remove all duplicate edges for each edge weight from $s_{p'}$ and keep edges unique by not deleting most recent edges for each edge weight.

Select elements from each disjoint graph as closed sequence and add it to DB_S and remove disjoint graphs from SG.

End C3
End: L3

If $SG \neq \phi$ go to L3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

VI. Comparative Study

This segment focuses mainly on providing evidence on asserting the claimed assumptions that 1) The PEPP is similar to BIDE which is actually a sealed series mining algorithm that is competent enough to momentously surpass results when evaluated against other algorithms such as CloSpan and spade. 2) Utilization of memory and momentum is rapid when compared to the CloSpan algorithm which is again analogous to BIDE. 3) There is the involvement of an enhanced occurrence and a probability reduction in the memory exploitation rate with the aid of the trait equivalent prognosis and also rim snipping of the PEPP. This is on the basis of the surveillance done which concludes that PEPP's implementation is far more noteworthy and important in contrast with the likes of BIDE, to be precise.

JAVA 1.6_ 20th build was employed for accomplishment of the PEPP and BIDE algorithms. A workstation equipped with core2duo processor, 2GB RAM and Windows XP installation was made use of for investigation of the algorithms. The parallel replica was deployed to attain the thread concept in JAVA.

Dataset Characteristics:

Pi is supposedly found to be a very opaque dataset, which assists in excavating enormous quantity of recurring clogged series with a profitably high threshold somewhere close to 90%. It also has a distinct element of being enclosed with 190 protein series and 21 divergent objects. Reviewing of serviceable legacy's consistency has been made use of by this dataset. Fig. 5 portrays an image depicting dataset series extent status.

In assessment with all the other regularly quoted forms like spade, prefixspan and CloSpan, BIDE has made its mark as a most preferable, superior and sealed example of mining copy, taking in view the detailed study of the factors mainly, memory consumption and runtime, judging with PEPP.

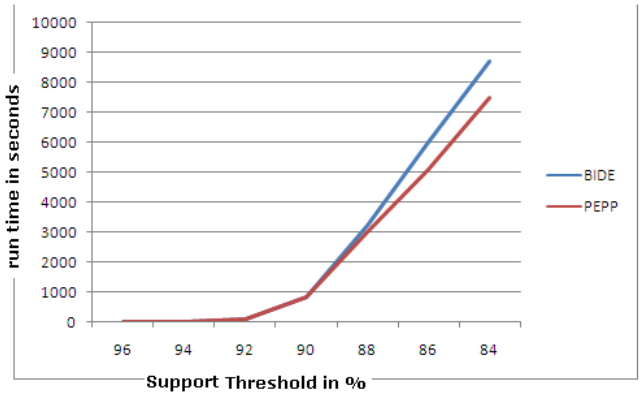


Figure 3: A comparison report for Runtime

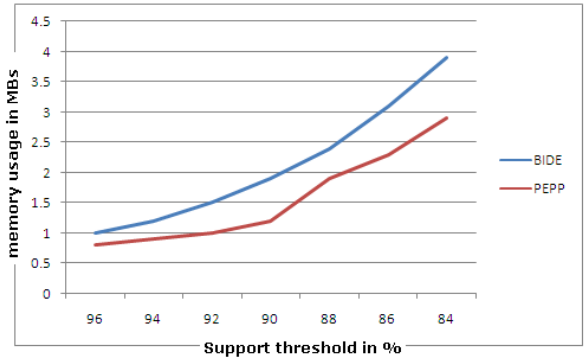


Figure4: A comparison report for memory usage

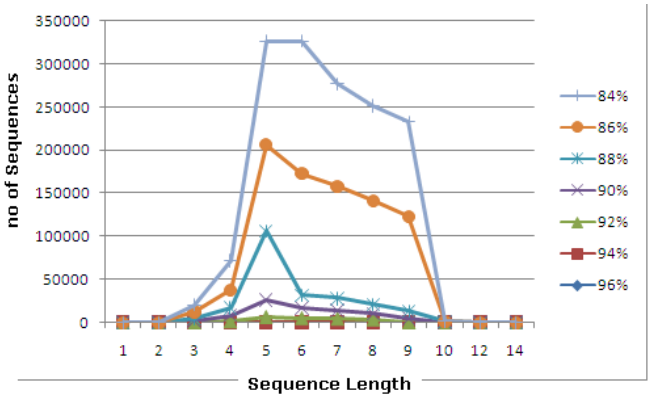


Figure 5: Sequence length and number of sequences at different thresholds in Pi dataset

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In contrast to PEPP and BIDE, a very intense dataset Pi is used which has petite recurrent closed series whose end to end distance is less than 10, even in the instance of high support amounting to around 90%. The diagrammatic representation displayed in Fig.3 explains that the above mentioned two algorithms execute in a similar fashion in case of support being 90% and above. But in situations when the support case is 88% and less, then the act of PEPP surpasses BIDE's routine. The disparity in memory exploitation of PEPP and BIDE can be clearly observed because of the consumption level of PEPP being low than that of BIDE.

VII. CONCLUSION

It has been scientifically and experimentally proved that clogged prototype mining propels dense product set and considerably enhanced competency as compared to recurrent prototype of mining even though both these types project similar animated power. The detailed study has verified that the case usually holds true when the count of recurrent moulds is considerably large and is the same with the recurrent bordered models as well. However, there is the downbeat in which the earlier formed clogged mining algorithms depend on chronological set of recurrent mining outlines. It is used to verify whether an innovative recurrent outline is blocked or else if it can nullify few previously mined blocked patterns. This leads to a situation where the memory utilization is considerably high but also leads to inadequacy of increasing seek out space for outline closure inspection. This paper anticipates an unusual algorithm for withdrawing recurring closed series with the help of Sequence Graph. It performs the following functions: It shuns the blight of contender's maintenance and test exemplar, supervises memory space expertly and ensures recurrent closure of clogging in a well-organized manner and at the same instant guzzling less amount of memory plot in comparison with the earlier developed mining algorithms. There is no necessity of preserving the already defined set of blocked recurrences, hence it very well balances the range of the count of frequent clogged models. A Sequence graph is embraced by PEPP and has the capability of harvesting the recurrent clogged pattern in an online approach. The efficacy of dataset drafts can be showcased by a wide-spread range of experimentation on a number of authentic datasets amassing varied allocation attributes. PEPP is rich in terms of velocity and memory spacing in comparison with the BIDE and CloSpan algorithms. ON the basis of the amount of progressions, linear scalability is provided. It has been proven and verified by many scientific research studies that limitations are crucial for a number of chronological outlined mining algorithms. Future studies include proposing of claiming a deduction advance on perking up the rule coherency on predictable itemsets.

REFERENCES

[1]F. Masseglia, F. Cathala, and P. Poncelet, The psp approach for mining sequential patterns. In PKDD'98, Nantes, France, Sept. 1995.

[2]R. Srikant, and R. Agrawal, Mining sequential patterns: Generalizations and performance improvements. In EDBT'96, Avignon, France, Mar. 1996.

[3]J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.C. Hsu, FreeSpan: Frequent pattern-projected sequential pattern mining . In SIGKDD'00, Boston, MA, Aug. 2000.

[4]M. Zaki, SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning, 42:31-60, Kluwer Academic Pulishers, 2001.

[5]J. Pei, J. Han, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.C. Hsu, PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. In ICDE'01, Heidelberg, Germany, April 2001.

[6]J. Ayres, J. Gehrke, T. Yiu, and J. Flannick, Sequential Pattern Mining using a Bitmap Representation. In SIGKDD'02, Edmonton, Canada, July 2002.

[7]M. Garofalakis, R. Rastogi, and K. Shim, SPIRIT: Sequential Pattern Mining with regular expression constraints. In VLDB'99, San Francisco, CA, Sept. 1999.

[8]J. Pei, J. Han, and W. Wang, Constraint-based sequential pattern mining in large databases. In CIKM'02, McLean, VA, Nov. 2002.

[9]M. Seno, G. Karypis, SLPMiner: An algorithm for finding frequent sequential patterns using length decreasing support constraint. In ICDM'02,, Maebashi, Japan, Dec. 2002.

[10]H. Mannila, H. Toivonen, and A.I. Verkamo, Discovering frequent episodes in sequences . In SIGKDD'95, Montreal, Canada, Aug. 1995.

[11]B. Ozden, S. Ramaswamy, and A. Silberschatz, Cyclic association rules. In ICDE'98, Olando, FL, Feb. 1998.

[12]C. Bettini, X. Wang, and S. Jajodia, Mining temporal relationals with multiple granularities in time sequences. Data Engineering Bulletin, 21(1):32-38, 1998.

[13]J. Han, G. Dong, and Y. Yin, Efficient mining of partial periodic patterns in time series database. In ICDE'99, Sydney, Australia, Mar. 1999.

[14]J. Yang, P.S. Yu, W. Wang and J. Han, Mining long sequential patterns in a noisy environment. In SIGMOD' 02, Madison, WI, June 2002.

[15]N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, Discovering frequent closed itemsets for association rules. In ICDT'99, Jerusalem, Israel, Jan. 1999.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[16]M. Zaki, and C. Hsiao, CHARM: An efficient algorithm for closed itemset mining. In SDM'02, Arlington, VA, April 2002.

[17]X. Yan, J. Han, and R. Afshar, CloSpan: Mining Closed Sequential Patterns in Large Databases. In SDM'03, San Francisco, CA, May 2003.

[18]J. Wang, J. Han, and J. Pei, CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Itemsets. In KDD'03, Washington, DC, Aug. 2003.

[19]R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In VLDB'94, Santiago, Chile, Sept. 1994.

[20]J. Pei, J. Han, and R. Mao, CLOSET: An efficient algorithm for mining frequent closed itemsets . In DMKD'01 workshop, Dallas, TX, May 2001.

[21]J. Han, J. Wang, Y. Lu, and P. Tzvetkov, Mining Top- K Frequent Closed Patterns without Minimum Support. In ICDM'02, Maebashi, Japan, Dec. 2002.

[22]P. Aloy, E. Querol, F.X. Aviles and M.J.E. Sternberg, Automated Structure-based Prediction of Functional Sites in Proteins: Applications to Assessing the Validity of Inheriting Protein Function From Homology in Genome Annotation and to Protein Docking. Journal of Molecular Biology, 311, 2002.

[23]R. Agrawal, and R. Srikant, Mining sequential patterns. In ICDE'95, Taipei, Taiwan, Mar. 1995.

[24]I. Jonassen, J.F. Collins, and D.G. Higgins, Finding flexible patterns in unaligned protein sequences. Protein Science, 4(8), 1995.

[25]R. Kohavi, C. Brodley, B. Frasca, L.Mason, and Z. Zheng, KDD-cup 2000 organizers' report: Peeling the Onion. SIGKDD Explorations, 2, 2000.

[26]Jianyong Wang, Jiawei Han: BIDE: Efficient Mining of Frequent Closed Sequences. ICDE 2004: 79-90

[27]Kalli Srinivasa Nageswara Prasad and Prof. S Ramakrishna. Article: Frequent Pattern Mining and Current State of the Art. International Journal of Computer Applications 26(7):33-39, July 2011. Published by Foundation of Computer Science, New York.

AUTHORS PROFILE:

Kalli Srinivasa Nageswara Prasad has completed M.Sc(Tech)., M.Sc., M.S (Software Systems)., P.G.D.C.S. He is currently pursuing Ph.D degree in the field of Data Mining at Sri Venkateswara University, Tirupathi, Andhra Pradesh State, India. He has published Five Research papers in International journals.

S.Ramakrishna is currently working as a professor in the Department of Computer Science, College of Commerce, Management & Computer Sciences in Sri Venkateswara university, Tirupathi, Andhra Pradesh State, India. He has completed M.Sc, M.Phil., Ph.D., M.Tech(IT). He is specialized in Fluid Dynamics and Theoretical Computer Science. His area of research includes Artificial Intelligence, Data Mining and Computer Networks. He has an experience of 25 years in Teaching Field. He has published 36 Research Papers in National & International Journals. He has also attended 13 National Conferences and 11 International Conferences. He has guided 15 Ph.D. Scholars and 17 M.Phil Scholars.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Hierarchical view for Level Set Method based on segmentation of Non- Constant Intensity Objects

M.Janani
M.Phil Scholar
P.S.G.R Krishnammal College For Women
Coimbatore-641004.

D.Kavitha Devi
Assistant Professor,
P.S.G.R Krishnammal College For Women
Coimbatore-641004.

***Abstract*—Segmentation of non-constant intensity object has been an important and vital issue for many applications. Segmentation of non- constant intensity object is a fundamental importance in image processing. Segmentation is difficult task in noisy images. The complementary method of the Mumford shah model for segmentation of non-constant intensity objects is been intended by level set method. The level set method retrieve the possible multiple membership of the pixels. Additive is forced through level set method which allows the user to control the degree of non-constant intensity objects and is more secure than the soft constraint the enhanced method increase efficiency, improve the effectiveness of segmentation. The numerical and qualitative analysis show that the level set algorithm provide more accurate segmentation result with good robustness.**

Keywords- level set method, non-constant intensity object, terzopoulos, kass, witkins, lipschitz.

I. INTRODUCTION

Segmentation is a process of dividing an image into meaningful, non-overlapping regions. Level set method is the process to improve the segmentation and simultaneously solving the non-constant intensity object. Segmentation of non-constant intensity object and incorporating some knowledge about their spatial relationship is a vital task. The problem of segmenting non-constant intensity object with possible occlusion in a variation setting is been solved. Hard segmentation model is that inherit the original property of the Mumford shah formulation to segment and smooth images in a coupled manner. Chan and Vese proposed a piece wise constant Mumford shah model in by further Mumford shah advances by using a level set formulation.

The Hard segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

In soft segmentation, there is the persistent control of the intensity. In the soft segmentation, the restraint is only loosely prosecuted. We call this model the soft segmentation. The soft segmentation reduces to the piecewise constant Mumford Shah segmentation model. The solution of the soft segmentation will approach to that of the hard segmentation.

The active of moving curves and surfaces, called the Level-Set Method. The level-set method is one computational technique for tracking a propagating interface over time, which in many problems has proven more accurate in handling topological complexities such as corners and cusps, and in handling complexities in the evolving interface such as entropy conditions and weak solutions. It is a robust scheme that is relatively easily to implement. Multiple regions are captured by a single contour demonstrating the topological transitions allowed by the models in level set implementation.

II. IMPROVED MUMFORD-SHAH MODEL

A. Mumford-shah model

The Mumford-shah model is one of the standard segmentation models. Mumford shah functional has been extensively used for image segmentation. Mumford shah algorithm obtains simultaneous functionality of both image smoothing and segmentation. The active contour is viewed as the set of discontinuities considered in the Mumford-shah formulation. The smooth estimate of the image is continuously estimated based on the current position of the curve.

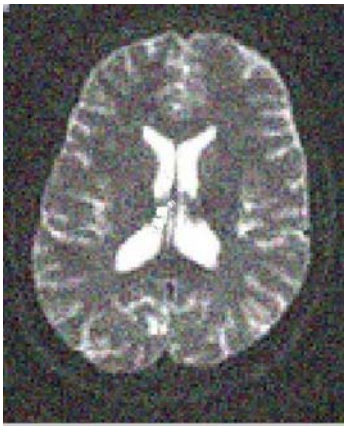


Fig. 1 Input image

Mumford-shah active contour model can handle image containing regions with roughly two different mean. Active contours were introduced by Kass, Witkins, and Terzopoulos for segmenting objects in images using dynamic curves. Munford-Shah model only can segment the image into two

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

parts according to the value of the shade of gray for specific images. The Mumford-shah model is not detecting the noisy image. The improved technique of Mumford-shah model is hard segmentation.

For a given image u_0, the piecewise constant Mumford-shah model seeks for a partition of Ω into N mutually exclusive open segmentsΩ_1,..... Ωn together with their interface C and a set of constant c=(c_1,c_2,….,c_n)which minimize the following energy functional:

$$F^{MS}(\Omega_1,\ldots,\Omega_n,c) = \sum_{i=1}^{n}\int_{\Omega_i} |u_0(x,y)-c_i|^2 dxdy + \alpha \times Length(C)$$

The idea is to partition the image so that the intensity of u_0 in each segment Ω_i is well-approximated by a constant c_i. The geometry of the partition is regularized by penalizing the total length of C. This increases the robustness to noise and avoids spurious segments.

B. *Hard mumford-shah model*

Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Given a fixed segmentation, it can be easily shown that the optimal constants are given by formulas denotes the Lebesgue measure of its argument Let us take non-constant intensity regions in that brain MRI image. Here, the over-line denotes the set closure. Although and generally do not constitute a partition, we still call the pair a segmentation of for simplicity. It should be clear that the partition is given by (together with the boundary of these segments inside).Given an image, the hard additive model seeks for a segmentation and a set of constants which minimize the energy, subject to an additive constraint .This model enforces a strict additive in the common region.

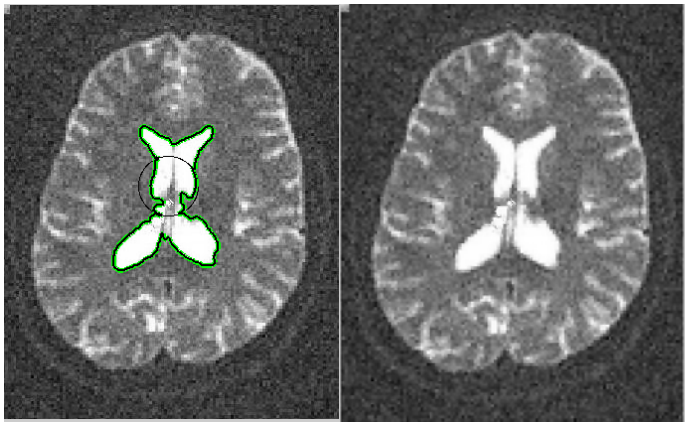


Fig.2 (a). 400 iteration of hard Mumford- shah model, (b). After noise removal in hard Mumford-shah model

Let Ω_1 and Ω_2 be two open regions in Ω that represent two computed objects. The following short hands to simplify the notations:

$$\Omega_{10} := \Omega_1 \setminus \overline{\Omega_2}$$
$$\Omega_{11} := \Omega_1 \cap \Omega_2$$
$$\Omega_{01} := \Omega_2 \setminus \overline{\Omega_1}$$
$$\Omega_{00} := \Omega \setminus \overline{\Omega_1 \cup \Omega_2}.$$

Here, the over-line denotes the set closure. Although Ω_1 and Ω_2 generally do not constitute a partition of Ω , we still call the pair {Ω_1 , Ω_2 }a segmentation of Ω for simplicity.

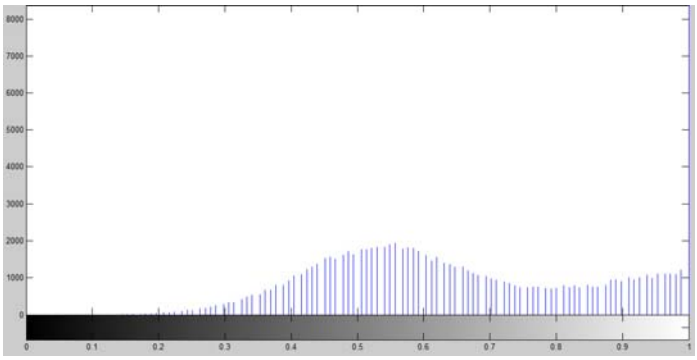


Fig.3 Histogram of hard Mumford-shah model

Given an image u_0 , the hard additive model seeks for segmentation {Ω_1 , Ω_2 } and a set of constants **c**=(c_{10},c_{01},c_{11},c_{oo}) which minimize the following energy:

$$\begin{bmatrix} c_{10} \\ c_{01} \end{bmatrix} = \begin{bmatrix} |\Omega_1| & |\Omega_{11}| \\ |\Omega_{11}| & |\Omega_2| \end{bmatrix}^{-1} \begin{bmatrix} \int_{\Omega_1} u_0 dxdy \\ \int_{\Omega_2} u_0 dxdy \end{bmatrix}$$
$$c_{00} = \frac{1}{|\Omega_{00}|}\int_{\Omega_{00}} u_0 dxdy.$$

Subject to an additive constraint c_{11}= c_{10}+c_{01}. Thus, this model enforces a strict additive in the common region.

C. *Soft mumford-shah model*

The soft segmentation reduces to the piecewise constant Mumford–Shah segmentation model. The solution of the soft segmentation will approach to that of the hard segmentation. Given segmentation, the optimal constants can be obtained by the formulas. The intensity level within each region has a certain degree of variation. A multi phase formulation with membership functions has recently been used with a different regularization term in for soft segmentation.

$$F^{soft}(\Omega_1,\Omega_2,\mathbf{c}) = \sum_{i=1}^{2}\int_{\partial\Omega_i\cap\Omega}[\alpha+\beta\phi(\kappa)]ds + \sum_{i=0}^{1}\sum_{j=0}^{1}\int_{\Omega_{ij}}(u_0(x,y)-c_{ij})^2 dxdy + \gamma(c_{10}+c_{01}-c_{11})^2$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Where $\gamma \geq 0$ is a constant controlling the degree of additive. In this model, $c_{10}+c_{01}=c_{11}$ the constraint is only loosely enforced. We call this model the soft additive model.

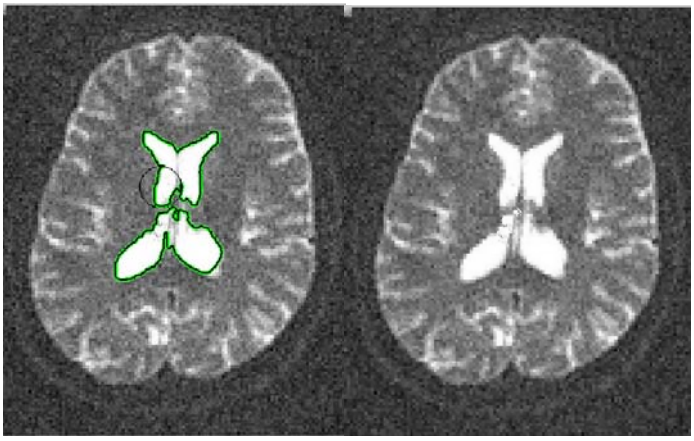


Fig.4 (a). 400 iteration of soft Mumford- shah model, (b). After noise removal in soft Mumford-shah model

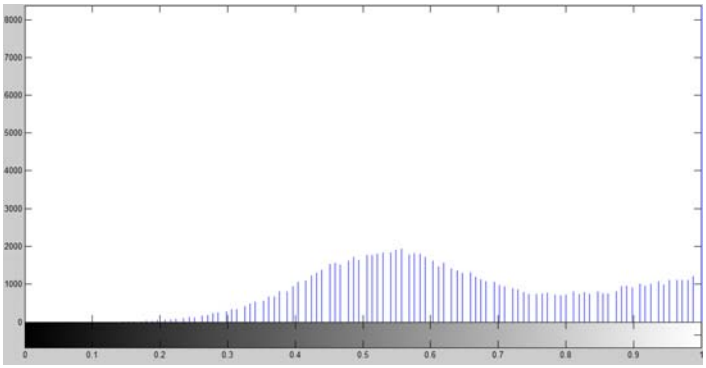


Fig.5 Histogram of soft Mumford-shah model

III. LEVEL SET METHOD

The level set method is a powerful tool which can be used for numerical realization. Level set representation is an established technique for image segmentation .Level set methods is to minimize a given function which aims to extract one or several elements of interest from the background. Level set method is referred to as a curve. In level set method, the curves are implicitly defined as the zeros of a lipschitz continuous function.

The level set method depended on the global information of homogeneity region, and is more robust than curve evolution model to detect discontinuities under noisy environment the level set method, can successfully handle the topology changes. Level set method has been applied to a variety of synthetic and medical images in different modalities. The level set method overcome the problem of soft segmentation problem it proves to be more accurate and robust. One way to represent a curve is as a level set or an equal-height contour of a given function. The Osher-Sethian level set formulation allows the development of efficient and stable numerical schemes in which topological changes of the propagating curve are automatically handled. The level set formulation puts curve evolution equation into the level set formulation. The level set method overcomes the problem of soft segmentation model .multiphase level set image segmentation. This method established on explicit correspondence between n region of segmentation and a partition defined using $\log_2 n$ level set functions.

Let $\omega I=\Omega C\ R^2 \rightarrow R$ IE [1,,,,,,M],be M level set function with $M=LOG_2N$.

Level set reprentation

$$if \begin{cases} \varphi(p) > 0 \ \ p\ belongs\ to\ the\ inner\ part\ \Omega_{in} \\ \varphi(p) < 0 \ \ p\ belongs\ to\ outer\ part\ \Omega_{out} \\ \varphi(p) = 0\ p\ is\ on\ the\ interface\ \Gamma \end{cases}$$

Where Ω_{in} is a region in Ω bounded by Γ,and Ω_{out} is defined as the complement of $\Omega_{in,}$ i.e. $\Omega_{out}=\Omega_{in.}$.To avoid unnecessary calculation and statistical errors the level set representation is used.

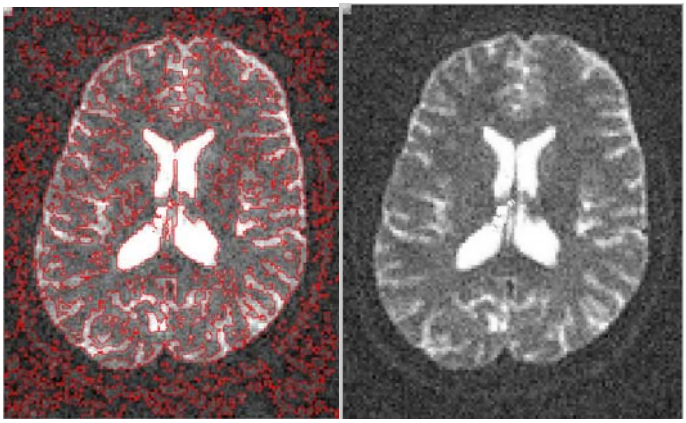


Fig.6 (a). 730 iteration of level set method, (b). After noise removal in level set method

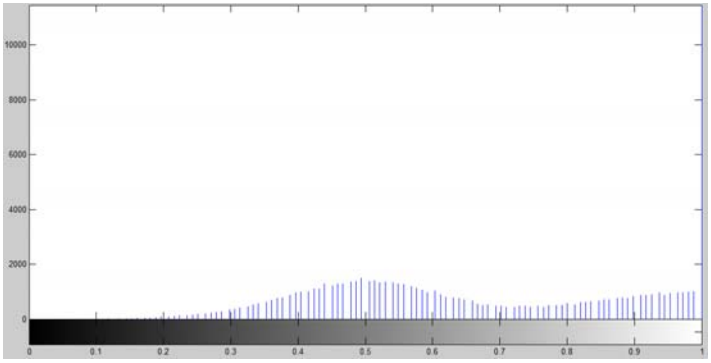


Fig.7 Histogram of level set method

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. EXPERIMENTAL RESULTS

The brain MRI image is used as an object in this paper. The brain MRI image was segmented using soft and hard Mumford-shah model; the segmenentation image is not accurate. So we validate our model based on; (a) performance comparisons between the hard segmentation, soft segmentation and level set method; (b) non-constant intensity object segmentation. In the hard segmentation the non-constant intensity objects is not been segmented accurately. In the soft segmentation the non-constant intensity objects is segmented but the output of the retrieval image is not accurate. In the level set method the non-constant intensity objects is been segmented.

Table.1 Intensity

Method	Original image	Overlapped image	Background image
Hard segmentation	101.0906	0.8208	0.4506
Soft segmentation	101.0906	0.8654	0.7973
Level set method	101.0906	0.9923	0.8112

Table.2 Standard deviation

Method	Original image	Overlapped image	Background image
Hard segmentation	101.0906	0.9291	0.5282
Soft segmentation	101.0906	0.9499	0.7453
Level set method	101.0906	0.9822	0.8541

CONCLUSION

The alternative method of the Mumford shah model for segmentation of non-constant intensity objects is been intended by level set method. The optimized zero level set indicate their approximate shapes and distributions clearly. Level set model has overcome some refractory challenges in elasticity reconstruction. The level set method is more robust than the soft segmentation with respect to global convergence. Hard segmentation fails to detect multiple non-constant intensity objects. The problem of segmenting non-constant intensity objects with possible occlusion in a variation setting is been solved. Level set method it solves the segmentation with depth problem that aims to recover the spatial order of non-constant intensity objects. Segmentation of multiple objects is been identified accurately. Finally, we demonstrate a hierarchical implementation of our model which leads to a fast and efficient algorithm capable of dealing with important image features.

REFERENCES

[1] J. Sokolowski and J.P. Zolesio, *Introduction to Shape Optimization*, Springer, New York, 1991.

[2] W. Zhu and T. Chan, "A variational model for capturing illusory contours using curvature," *J. Math. Imag. Vis.*, vol. 27, pp. 29–40, 2007.

[3] T. Chan and L. Vese, Active contours without edges, *IEEE Trans. On image proc.*, 10(2):266–277, Feb. 2001.

[4] C.A.Cocosco, A.P.Zijdenbos, A.C.Evan, "A Fully Automatic and Robust Brain MRI Tissue Classification Method," *Medical image Analysis,* 7, pp.513-527, 2003.

[5] A.Chakraborty, L.H.Staib, and J.S.Duncan, "Deformable Boundary Finding in Medical Images by Integrating Gradient and Region Information," *IEEE Trans. On Med. Imag.*, 15(6) pp. 859-870, 1996.

[6] V.Caselles, R.Kimmel, and G.Sapiro, "Geodesic Active Contours," *Int. Journal of Computer Vision.* 22(2), pp. 61-79, 1997.

[7] Faugeras O. and Keriven R., Variation principles, surface evolution and PDE's level set methods, and the stereo problem. *IEEE Trans image processing.,* 1998, 7(3):336-344.

[8] J. A. Sethian, Level set methods and fast marching methods, Cambridge: Cambridge University Press, 1999.

[9] T Chan and L. Vese, "Active Contour without Edges", *IEEE Transaction on Image Processing,* 2001, 10(2), pp.266-277.

[10] D. Mumford. J. shah. "Boundary detection by minimizing functional" Proc. of IEEE CVPR 1985.

[11] S. Osher. J. Sethian, "Fronts propagating with cur\aturedependent speed: algorithms based on Hamilton-Jacobi Formulation", *Journal of Computatzonal Physzcs.* 79. 12-49. 1988.

[12] Li-Tien Cheng, Paul Burchard, Barry Merriman, and Stanley Osher, "Motion of curves constrained on surfaces using a level set approach". *J. Comput. Phys.*, 175:604.644, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Customer Relationship Management and its Implementation in E-Commerce

Mohammad Behrouzian Nejad
Young Researchers Club, Dezfoul Branch
Islamic Azad University, Dezfoul, Iran
Mohamad.behrouzian@gmail.com

***Abstract*—In the new business gain customers is dedicated to its crucial position in the organization's goals and senior managers are well aware that their success in achieving the overall goals of the organization depends on satisfying our customers. Customer Relationship Management (CRM) includes all the steps that an organization to create and establish beneficial relationships with the customer takes. CRM is now a core is allocated to the business world. E-commerce, one of the issues raised the effect of information technology is expanding. Regarding the topic of CRM is important in E-Commerce and management, this paper shows the impact of CRM to improve relationships with customers in organizations and E-Commerce. Also, examines the general definition of CRM, goals, benefits, success factors and implementation of traditional and Electronic CRM.**

Keywords-Customer Relationship Management, E-Commerce, Information Technology, Organizations.

I. INTRODUCTION

Today customers play key roles in the development and direction of the play activities of any organization. Organizations have found that understanding the needs and create value for customers the main factor for success in a competitive world. Business culture has made progress in recent years and consequently economic relations and the fundamental approach to customers are changing. Technological change and intense competitive environment has made conditions difficult for manufacturers and service providers that can not be found for longer years high demand and stable or guaranteed markets were steady customers. So in today's world using systems such as CRM is not only a competitive advantage, rather it is considered a necessity for organizations. One of the systems can have significant influence on organizational decisions is CRM (CRM). CRM pays to gather information based on current and future needs and demands of customers. CRM offers comprehensive information about the supply chain management that help is on decision making in order to estimate a step closer to customer needs and designing an organization around customers. In fact CRM is an important issue in today's global economy that will force the organization to rethink the strategy for communicating with customers and capture a wide range of knowledge and identify loyal customers [1-7].

II. CRM

Relationship management is a strategy for selecting and managing customers in order to create value in the long run. CRM System is a business strategy through software and techniques that have been linked to helping to manage more effectively communicate with customers in direct or indirect channels. CRM uses from one to one marketing to customize the product to the customer which is a continuous process of data collection in all the time with the customer then converts this data into knowledge to communicate more effectively with customers in order to be more profitable. A key to success in CRM not having a lot of customer data, but how important is their use by companies [8]. Another definition of CRM in [2]: Organizational approach to understanding and influencing customer behavior through meaningful communication in order to attract customers, customer retention, customer loyalty and customer profitability. In fact CRM is a strategy working with this approach which with customers commensurate with qualifications and their behavioral patterns established sustainable and long-term relationship that added value for both sides. CRM strategy is usually based on a four-goal is the implementation:

- Encourage customers of other companies or potential customers to first purchase of the company.
- Encourage customers who first made the purchase to next purchases.
- Conversion of temporary customers to loyal customers.
- Provide high quality services for loyal customers, so that will be converted to the advertiser for the company.

In fact CRM including all processes and technology that the organization is working to identify, select, promote, develop, maintain and service the customer. CRM will enable managers that use of customer knowledge to boost sales, service and its development [4].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III. CRM HISTORY

Perhaps can be summarized the history of topics related to CRM in the following three period [2, 8]:

A. Handicraft production to mass production (during the Industrial Revolution)

Ford's initiative in employing methods of mass production to replace manual methods of production, this is one of the most important indicators. However changes in production practices led to the selection of customers, in terms of product characteristics is reduced (compared to the first category), but the products of the new method the price was lower on the Curb. In other words the chosen method of mass production the Ford, increase efficiency and cost effectiveness the most important goals were predicted.

B. Mass production to continuous improvement (during the Quality Revolution)

This period began with the initiative of the Japanese company's continuous improvement process. This in turn led to low production costs and higher quality products. This course is being introduced with new methods of quality management such as TQM, reached its peak. But with the increasing number of companies in the competitive arena and the culture to maintain and improve product quality (through various quality means), another competitive advantage for companies did not leading and work and will feel the necessity of finding new ways to maintain a competitive advantage.

C. Continuous improvement to mass customization (during the Customer Revolution)

In this period due to the increasing expectations of customers, manufacturers were forced to produce their products with low cost, high quality and high diversity. Other words, manufacturers were forced to focus their production, only to find ways to satisfy and retain their previous customers have paid.

IV. CUSTOMER TYPES

Can be viewed to the customer categories from two different perspectives, the first category deals with customer from the perspective of his place [8]:

- Foreign customers: refers to a collection which receives services from Organization.
- Customer middle: refers to a collection, which receive services from Organization and with change or no change, it will be delivered to foreign customers.
- Internal customers: the new wave of management, which refers to all employees of an organization that will have to step in for providing services.

But the other view, expressed to the same classification in terms of customer satisfaction and classification of the customer, in terms of utility service and his loyalty to the organization, completely loyal to the level of opposition. The CRM perspective, the view outside with external customers and internal customers or internal staff's attitude and the second point is converting customers with different levels of satisfaction to very loyal customers. Gandhi, the late leader and the people of India, the key years of his life sentences without the knowledge of the issues of customer orientation has been expressed which can be useful. Part of it is the following:

- Customer is most important supervisor on our activities.
- He is not dependent on us.
- We are dependent on him.
- Customer in our work is not a fleeting purpose.
- His ultimate aim of all our actions.
- Customer for we is not a foreign one.
- He is our part.
- We with serve customers do not favor but they give us a chance to work.
- So we should thank him.

Also perhaps, four indicators of customer orientation is seen as the main components, which include:

- Measuring and understanding customer needs in order to satisfy him.
- Preparation for the changing needs and requirements change.
- Attempt to provide impeccable service.
- Customer orientation is an issue in the organization, for all categories, all duties and roles.

V. RELATIONSHIP BETWEEN IT AND CRM

Traditional marketing does not require much use of information technology, because there is no need to identify, isolate and distinguish between customers, customer interaction and customization of customer needs. While these four functions in the CRM, a lot depends on their technology and information systems. It should be noted that the strategic CRM which to help them, we learn more about customers' needs and behaviors and relationships stronger, friendlier and more useful to have with them. In fact, having a good relationship with customer is heart of any successful and healthy business. However, it is wrong which know CRM as a technological phenomenon. Rather, CRM is essentially a process and IT can have role of facilitate this process. The main idea lies in the integration of CRM and information technology, combining skills in information technology and human resources to achieve a deep insight about the customer's wishes and values. CRM should be able to meet the following objectives in this way [9]:

- Providing better services for customers.
- Raising productivity of telephone company facilities.
- Help to sales personnel to expedite transactions.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Simplify marketing and sales processes.
- Discover new customers.
- Raising the income level of customers.
- Increase customer satisfaction and enhance their loyalty to the company and products.

VI. E-COMMERCE

Definition of e-commerce, according to various definitions and concepts in the CRM, this is the case: use of networks and related technology to automate, improve, upgrade or complete redesign business systems to create more value for customers and business partners. The Internet has opened a new arena for dissemination, exchange and present information that is placed facing humanity in many ways is a profound revolution. The revolution to this concept, which foundations of economic, social, cultural, political and technological communities gradually will change. In the near future will be conducted the volume of exchanges of scientific, educational, economic, marketing, tourism and many community activities, exclusively via the Internet. In a sentence can be said that all roads will lead to the Internet. Electronic- Commerce has been one of the innovative using of Internet in business. E-Commerce is a phenomenon that is growing which has attracted many different businesses. In general term can be stated that E-Commerce involves using electronic means to exchange goods, services and information and wide variety of electronic tools that are used for this purpose which can be referred to the internet, intranet, extranet and ... as a means of communication in e-commerce. Currently, Internet is the most common tool used in e-commerce. The main activities of suppliers and buyers in e-commerce, coordinating their work activities with the Internet and learn how to manage the business on the Internet, E-Commerce is the main challenge. From E-Commerce benefits for its participants can be mentioned this cases [10]:

- Reducing costs.
- Increasing efficiency which increases speed and accuracy in performing the tasks.
- Access to wider markets and specific markets for a particular product or service.
- Access to products with lower prices and higher quality.
- Transaction easy and accurate access to information.
- Access to a wide range of products and... .

VII. ELECTRONI CRM

Electronic Customer Relationship Management (E-CRM) is a marketing strategy, sales and online services which are integrated which can be play a role in identify, acquire and retain customers that are the largest investment companies. Electronic CRM will improve and increase communication between company and customers by create and enhance relationships with customers through new technology. Electronic CRM software, provide profiles and a history of contact with the any customers. Electronic CRM is a combination of hardware, software, applications and management obligations. Daycheh noted there are two types of E-CRM: Operational E-CRM and Analytical E-CRM. Operational E-CRM including customer contact centers such as telephone, fax and e-mail that Company has been in contact with customers this way and is includes marketing and sales which is done by special teams. Analytical E-CRM, technology needs to provide large amounts of data from the customer, which purpose of this section is to analyze customer data, purchasing patterns and other important factors, which will create new business opportunities. E-CRM according to the organization, takes on different forms. E-CRM is not only software and technology, but also includes business processes based on customer-centric strategy that is supported by various software and technology [8].

VIII. PERFORMANCE OF E-CRM

In today's world, organizations are communicating with customers through various communication channels such as World Wide Web, call centers, finding a market, the vendors and partners. E-CRM systems will encourage customers to do business with the organization and provides a way which in it customers can receive any type of product at any time from any channel and with any language that would be and because with they treated as a unique person, they feel comfortable. E-CRM systems provide a central repository for recording and storing information about customers and put it in the computer system employees and each employee can have access to customer information at any time. Benefits of E-CRM including [8]:

A. Increased customer loyalty

E-CRM system allows the company, despite the various communication channels to communicate with customers via both individual and unique. Using E-CRM software, anyone can achieve that in the organizations history and customer information. Information obtained by using E-CRM helps to company that meet the cost of obtaining a true Holly and maintain customer individually. Having this data allows companies to focus time and resources more beneficial to the customers. This tool creates the possibility that the customer for each purchase will go to the website, the company's customers know and according to the profile of the customer, facilitate the process of shopping for her. Using customer profiles in CRM, it has been used in pointing out the customers who are loyal [1, 8].

B. Efficient Marketing

Having customer information by an E-CRM system allows the company to predict a variety of products to customers interested in buying them. This information helps to put the organization, its marketing and sales, with more efficiency and effectiveness in order to satisfy the customer. Customer data from different perspectives to create the right marketing for more useful products are analyzed. Another benefit of segmenting customers, improve marketing process. Segmenting customers according to they needs, allows to company private products to customers.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

C. Improve to Efficient Services and Support

An E-CRM system provides a single repository of customer information. This work enables companies to at all customer contact centers, perform the customer needs with speed and high efficiency. E-CRM technologies are includes: search engines, live and online help, e-mail management, news management and supporting different languages. With an E-CRM system a company can:

- Orders received with complete accuracy, update and run.
- Record information, costs and time associated with ordering information.
- See the customer service contracts.
- Search is the most reliable and best practice solutions.
- Be a member, information sites and product-centric and software.
- Access to knowledge tools, which are useful in order to complete the service.

D. Higher Efficiency and Lower Costs

Using various techniques such as data mining that is data analysis, is know as relationship between parts of data, can be a valuable resource. Customer information is collected in a single database to all components within the company (marketing team, sales force, etc.) allows this to can to share together its information and jobs [8].

IX. Importance and Benefits of CRM

Retaining customers in all industries especially small and medium industries according to limited resources is important. In addition dissatisfied customers, it makes the organization vulnerable to market. Because they do harm to competition and other customers also convinced that it would avoid the transaction with the organization. It is clear that CRM is an important issue in the debate business world. Primary researchers of CRM think's interests of CRM, structure of any industry were searched separately. But the results of recent investigations in several countries and has conducted several industry shows that the interests CRM in different industries and countries, does not change much. The main advantages of CRM including:

- Improved capabilities in targeting profitable customers.
- Virtual integration of communication with customers.
- Improved sales force efficiency and effectiveness.
- Personalized marketing message.
- Proportional to (especially storage) products and services.
- Improved efficiency and effectiveness in customer services.
- Improved pricing.

X. Factors of Success in CRM

In the analysis of CRM is cited as success factors. Although the implementation of CRM in marketing and sales service organization has been redesigned but success in the cultural field, as there is non-technical factors [11, 12].

- *Evolution*: in order, step by step implementation of CRM is from the practical and analytical to the cooperation and coordination. For example, many companies in the practical phase of CRM using of sales force or call centers.
- *Time Efficiency*: a complete system in a preliminary stage lasts about 7 months. At this stage for complete database with meaningful data, in the sales marketing and services should use the information that is at least about 2 years.
- *Organizational Redesign*: Create a center of the accountability and defining standards to avoid cultural conflict.
- *Change management.*
- *Senior management support.*

XI. Implementation of CRM in E-Commerce [13]

A. Clear definitions of groups of customers

Our customers are not only those of us who are buying, but people who are interested in our company and make suggestions and give them as *lies Customers* is remembered, too are our customers. The company must be using a variety of methods these customers will become the real secret to our customers.

B. Complete category management market

In E-Commerce to accommodate CRM and enterprise resource planning and supply chain management, there is only one way and it is complete category management market.

C. Establish communication channels with all kinds of customers

With establish communications especially direct contact with customers can raise customer satisfaction of services which with this Work, customers build for themselves ideal image of the company.

D. Correct idea of management

Managers should have long-term supervision. They must know their customers as the company's wealth and capital. Managers must accept that the leaders of the organization and organizational change start from leadership point. They should be in line with the idea that *our customers are important*, create CRM systems and this process, but with the advantage of E-Commerce through networks and empowering employees to satisfy customers, does not exist.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

XII. CONCLUSION

Organizations in implementing a customer-centric strategy for customers must with create incentives in between customers and their employees lead them by cooperate with each other, in order to finalize the company's strategy. Today increasing access to information changes in supply and demand has shifted power from the seller to the customer. Therefore preferable is created through personal interaction with customers and understand customer needs. CRM system is a system which helps to organization to maintain customer's long-term relationship with the organization. With the advent of Internet and E-Commerce development how trade and exchange has taken a new shape. Many organizations to reduce vulnerability in relation to customers are under implementation or planning to implement CRM systems. So many organizations and corporations, the projects have been implemented to make progress in the field of customer orientation and be able to plan and implement CRM systems. Customer-oriented management requires having the appropriate technical infrastructure, economic and human resources. Obviously for e-business development, entry into global markets and membership in organizations such as WTO, CRM is among the basic requirements. Among the expected benefits of CRM can be pointed to increased customer satisfaction, customer care to create products, services and special and differentiated value for customers.

REFERENCES

[1] *P. S. Vadivu, V. K. David, "*Enchancing And Deriving Actionable Knowledge From Decision Trees*"*, International Journal of Computer Science and Information Security, Vol. 8 No. 9, pp 230-236, 2010.

[2] A. Ebrahimi, H. Farzad, "Performance of Analytical CRM for Credit Scoring of Bank Customers Using Data Mining Algorithms", Proceedings of 2nd National Conference on Information Technology, Present, Feature, Islamic Azad University, Mashhad Branch, Iran, 2010.

[3] A. Ansariasl, A. Einipoor, M. Nikan, "Using new techniques of knowledge management to improve Customer Relationship Management in firm business", Proceedings of 1st Regional Conference on Management of Young Researchers Club, Islamic Azad University, Shoshtar Branch, Iran, 2010.

[4] K. J. Firoozabadi, E. Darandeh, S. B. Ebrahimi, "Presentation technique of customer knowledge management based on simultaneous establishment of KM and CRM in the organization", Proceedings of Industrial Engineering and Management, http:\\www.betsa.ir, 07/04/2011.

[5] M. Behrouzian-Nejad, E. Behrouzian-Nejad, H. Shayan, N. Mousavianfar, "role of Customer Relationship Management in organizations", Proceedings of 1st Regional Conference on Management of Young Researchers Club, Islamic Azad University, Shoshtar Branch, Iran, 2010.

[6] A. Afrazeh, "Provided a model for established relationship between quality management and information quality management in supply chain", Proceedings of 1st National Conference on Logistics and supply chain, Industrial University of Amirkabir, pp 1- 11, Iran, 2004.

[7] A. Ashkari, E. Akhondi, F. Fatholahi, S. Sayadmanesh, "Integration of supply chain management and Customer Relationship Management", Proceedings of 1st National Conference on Logistics and supply chain, Tehran, Iran, 2006.

[8] S. Emamgholizadeh, F. Chaman, "Survey of applying method of electronic communication with the customer in organization", Proceedings of 1st Regional Conference on Management of Young Researchers Club, Islamic Azad University, Shoshtar Branch, Iran, 2010.

[9] R. Roozbahani, M. Modiriasari, "Technology of Customer Relationship Management", http:\\www.hajarian.com, 07/04/2011.

[10] M. Akhavan, M. Babolhavaegi, " Customer Relationship Management in e-commerce between the firm", Proceedings of 4th International Conference on Industrial Engineering, Tarbiat Modares University of Tehran, Iran, 2005.

[11] V. Ramezanizadeh, "Customer Relationship Management", http:\\www.bih.ir, 07/04/2011.

[12] R. Ling, D. C. Yen, "*Customer Relationship Management: An analysis framework and implementation strategies*". Journal of Computer Information Systems, 41,pp 82–97, 2001.

[13] Jibin, Ma, Sun Yonghao, Wu Xuyan & Chen Xiaoyan, *"Research of Customer Relationship Management in Enterprise under the E-Commerce"*,Vol 5, No 09, Hebei University of Engineering, pp. 131-134, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Decision Tree Based Model to Identify the Career Focus of Computer Stream Students in ITES Industry

T.Hemalatha[#1], Dr.Ananthi Sheshasaayee[#2]

#1. T.Hemalatha, Research Scholar, R&D centre, Bharathiar University,Combatore, Asst.Prof, M.C.A.Dept,VELS University,Chennai,India

#2. Dr.Ananthi Sheshasaayee- Associate Professor and Head, Dept of Computer Science, Quaid-E-Millath Govt. College for women (Autonomous), Chennai –600 002,India

[1]winhema18@yahoo.co.in

[2]ananthiseshu@gmail.com

Abstract - This paper focuses on the various career opportunities that are available for the computer stream students in the field of ITES industry .This paper analyses the various attributes of the Skill Set of computer stream students, from which a decision tree can be generated to help them to improve the confidence level of students in selecting a career in ITES industry. For the past few years it has become a passion for students to choose computer science as their main stream for their studies. During the final semester of their graduation they struggle a lot to choose a career based on the skill set they posses which is of due importance. With the use of Decision tree this paper provides a guideline to take decision to choose career in ITES Industry.

Keywords - skill set, career, computer stream, ITES, decision tree, decision

I. INTRODUCTION

In this competitive world it is very difficult to secure a job .Students who have chosen computer as their main stream can decide their career based on the skill set they posses. They are not much aware about the skills required by the ITES industry. Knowing the skills they possess, we can give them a decision to choose their career.

In our day today life, we come across various decision making problems. Normally we solve these problems and make decisions out of the experience which may be incorrect seldom. The computer technology helps us to provide an easy and efficient way of decision making. One such approach is the decision tree ,which is utilized in this paper. Decision tree learning is one of the most successful learning algorithms, for its various attractive features. Simplicity, comprehensibility, parameter less, and being able to handle mixed type data. In decision tree learning, a decision tree is induced from a set of labeled training instances represented by a tuple of attribute values and a class label. Because of the vast search space, decision tree learning is typically a greedy,top-down and recursive process starting with the entire training data and an empty tree. An attribute that best partitioned into disjoint subsets satisfying the values of the splitting attribute, for each subset, the algorithm proceeds recursively until all instances in a subset belong to the same class [1].

Decision trees are a rapid and effective method of classifying data set entries, and can offer good decision support capabilities. A decision tree is a tree in which each non-leaf node denotes a test on an attribute of cases, each branch corresponds to an outcome of the test, and each leaf node denotes a class prediction. The quality of a decision tree depends on both its classification accuracy and its size.[2] Existing studies have identified several advantages to the use of decision trees: no domain knowledge is needed for classification, they are able to handle high dimensional data, they are intuitive and generally easy to comprehend, they are simple and fast, and they have good accuracy [3].

II. RESEARCH METHODS

1. Data collection

Data for this study were collected from various ITES workers like developers, web designers, System admin, Team lead, Project Manager, developer, Testers, copy editor, reference setting, Hr, network admin, etc. We contacted many workers from ITES industry and a questionnaire was given to them The questionnaire is of closed type. Participation in this study was voluntary and people were assured that their individual responses would be treated as confidential.

2. Data Set Description

Among the various attributes the following attributes are considered as vital for career decision in ITES Industry. The vital attributes are 1. Communication skill 2. Knowledge on productivity software 3. Domain Knowledge 4. Soft Skill 5 Decision making 6. Analytical skills..

III. EXPERIMENTAL RESULTS

Based on the answer given by the them, the important attribute is selected on the basis of highest percentage and then attribute was categorized by the requirement needed by the ITES industry. It has been identified that the major skills set which are required for BPO companies are Knowledge on productivity software, Communication skill, Domain Knowledge, Soft Skill and for KPO companies, Knowledge on productivity software, Communication skill, Domain Knowledge, Soft Skill, Decision making, Analytical skills are required. We Assign Knowledge on productivity software as KPS, communication skill as CS, Domain Knowledge as DK, Soft skill as SS, Decision Making as DM and Analytical skill as AS.

Table I. Data Set for the ITES Industry

SAMPLE	CS	KPS	DK	SS	DM	AS	BPOEF	KPOEF
P1	1	1	1	1	1	1	1	1
P2	1	1	1	1	1	0	1	0.8
P3	1	1	1	1	0	1	1	0.8
P4	1	1	1	1	0	0	1	0.7
P5	1	1	1	0	1	1	0.8	0.8
P6	1	1	1	0	1	0	0.8	0.7
P7	1	1	1	0	0	1	0.8	0.7
P8	1	1	1	0	0	0	0.8	0.5
P9	1	1	0	1	1	1	0.8	0.8
P10	1	1	0	1	1	0	0.8	0.7
P11	1	1	0	1	0	1	0.8	0.7
P12	1	1	0	1	0	0	0.8	0.5
P13	1	1	0	0	1	1	0.5	0.7
P14	1	1	0	0	1	0	0.5	0.5
P15	1	1	0	0	0	1	0.5	0.5
P16	1	1	0	0	0	0	0.5	0.3
P17	1	0	1	1	1	1	0.8	0.8
P18	1	0	1	1	1	0	0.8	0.7
P19	1	0	1	1	0	1	0.8	0.7
P20	1	0	1	1	0	0	0.8	0.5
P21	1	0	1	0	1	1	0.5	0.7
P22	1	0	1	0	1	0	0.5	0.5
P23	1	0	1	0	0	1	0.5	0.5
P24	1	0	1	0	0	0	0.5	0.3
P25	1	0	0	1	1	1	0.5	0.7
P26	1	0	0	1	1	0	0.5	0.5
P27	1	0	0	1	0	1	0.5	0.5

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

P28	1	0	0	1	0	0	0.5	0.3
P29	1	0	0	0	1	1	0.3	0.5
P30	1	0	0	0	1	0	0.3	0.3
P31	1	0	0	0	0	1	0.3	0.3
P32	1	0	0	0	0	0	0.3	0.2
P33	0	1	1	1	1	1	0.8	0.8
P34	0	1	1	1	1	0	0.8	0.7
P35	0	1	1	1	0	1	0.8	0.7
P36	0	1	1	1	0	0	0.8	0.5
P37	0	1	1	0	1	1	0.5	0.7
P38	0	1	1	0	1	0	0.5	0.5
P39	0	1	1	0	0	1	0.5	0.5
P40	0	1	1	0	0	0	0.5	0.3
P41	0	1	0	1	1	1	0.5	0.7
P42	0	1	0	1	1	0	0.5	0.5
P43	0	1	0	1	0	1	0.5	0.5
P44	0	1	0	1	0	0	0.5	0.3
P45	0	1	0	0	1	1	0.3	0.5
P46	0	1	0	0	1	0	0.3	0.3
P47	0	1	0	0	0	1	0.3	0.3
P48	0	1	0	0	0	0	0.3	0.2
P49	0	0	1	1	1	1	0.5	0.7
P50	0	0	1	1	1	0	0.5	0.5
P51	0	0	1	1	0	1	0.5	0.5
P52	0	0	1	1	0	0	0.5	0.3
P53	0	0	1	0	1	1	0.3	0.5
P54	0	0	1	0	1	0	0.3	0.3
P55	0	0	1	0	0	1	0.3	0.3
P56	0	0	1	0	0	0	0.3	0.2
P57	0	0	0	1	1	1	0.3	0.5
P58	0	0	0	1	1	0	0.3	0.3
P59	0	0	0	1	0	1	0.3	0.3
P60	0	0	0	1	0	0	0.3	0.2
P61	0	0	0	0	1	1	0.0	0.3
P62	0	0	0	0	1	0	0.0	0.2
P63	0	0	0	0	0	1	0.0	0.2
P64	0	0	0	0	0	0	0.0	0.0

In the Table I. strong skill set value is represented as 1 and the weak skill set value is represented as 0.The value ranges for BPO Key Factor and KPO Key factor varies from 0.0 to 1.0.

BPO Key Factor {0.0 – 1.0}

KPO Key Factor {0.0 -1.0}

BPO Eligibility Factor is represented as (BPO EF) and KPO Eligibility Factor is represented as (KPO EF) .SOSS1 represents sum of skill set of BPO EF. For each attribute the values assigned to them is 1.

SOSS1= Summation of the four attributes they are communication skills, knowledge on productivity software, domain knowledge, soft skills .ie.SOSS1= $\sum$ [(cs) +(kps) +(dk)+(ss)

SOSS2= Summation of the all six attributes they are communication skills, knowledge on productivity software, domain knowledge, Soft Skill ,decision making, Analytical skills i.e. $\sum$ ((cs) +(kps) +(dk)+(ss) +(dm) +(as))

BPO Attribute Factor is represented as BPO AF and its value is 4. {CS,KPS,DK,SS}

KPO Attribute Factor is represented as KPO AF and its value is 6.{CS,KPS,DK,SS,DM,AS}

BPO EF=SOSS1/BPO AF

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

KPO EF=SOSS2/KPO AF
The values of BPO EF and KPO EF can be illustrated through decision tree as depicted below.

Decision trees
A decision tree is a flow-chart-like tree structure allowing to determine the class of an
object given known values of its attributes. The visual presentation makes the decision tree
model very easy to understand. It is composed of three basic elements:
1. A decision node specifying the test attribute.
2. An edge corresponding to one of the possible values of the test attribute outcomes. It leads generally to a sub decision tree.
3. A leaf which is also named an answer node, including objects that, typically, belong to the same class, or at least are very similar. For what concerns a decision tree, the developer must explain how the tree is constructed and how it is used:
• Building the tree: Based on a given training set, a decision tree is built. It consists in selecting for each decision node the appropriate test attribute and also to define the class labeling each leaf.
• Classification: Once the tree is constructed, it is used in order to classify a new instance.
We start at the root of the decision tree, we test the attribute specified by this node. The
result of this test allows us to remove down the tree branch according to the attribute
value of the given instance. This process is repeated until a leaf is encountered and
which is characterized by a class.[4]

In many real-world problems, classes of examples in the training set may be partially defined and even missing.For example, for some instances, an expert may be unable to give the exact class value. A doctor who cannot specify the exact disease of a patient, a banker who cannot decide whether to give or not a loan for a client, a network administrator who is not able to decide about the exact signature of a given connection, etc. Hence, in these different examples, the expert can provide imprecise or uncertain classifications expressed in the form of a ranking on the possible classes. Ignoring the uncertainty may affect the classification results and even produce erroneous decisions. Consequently, ordinary classification techniques such as decision trees should be adequately adapted to take care of this problem[5]Decision tree can handle big amounts of data. Their representation of acquired knowledge in tree form is intuitive and generally easy to assimilate by humans [6].Decision tree is popular tool in data mining , it is also well suited for the classification task in that it is reasonably good at a variety of classification task.[7,8]

In Fig 2. The value 2 represent the same branch of KPS. Fig1. Shows the decision tree for the BPO industry and Fig 2. Shows the decision tree for the KPO industry. By knowing the skill set we can able to identify the chances of getting into ITES industy.In the Decision Tree first we have to check the first skill set ,if he posses the skill set then the condition is yes and we have to check the next skill set. If he does not possess the required skill then the condition become no then he have to improve(IM) that skill and we can check for the next skill .Likewise we have to check all the skill set and find out the eligibility factor. By finding the eligibility factor we can able to say the chances of getting a job in ITES

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

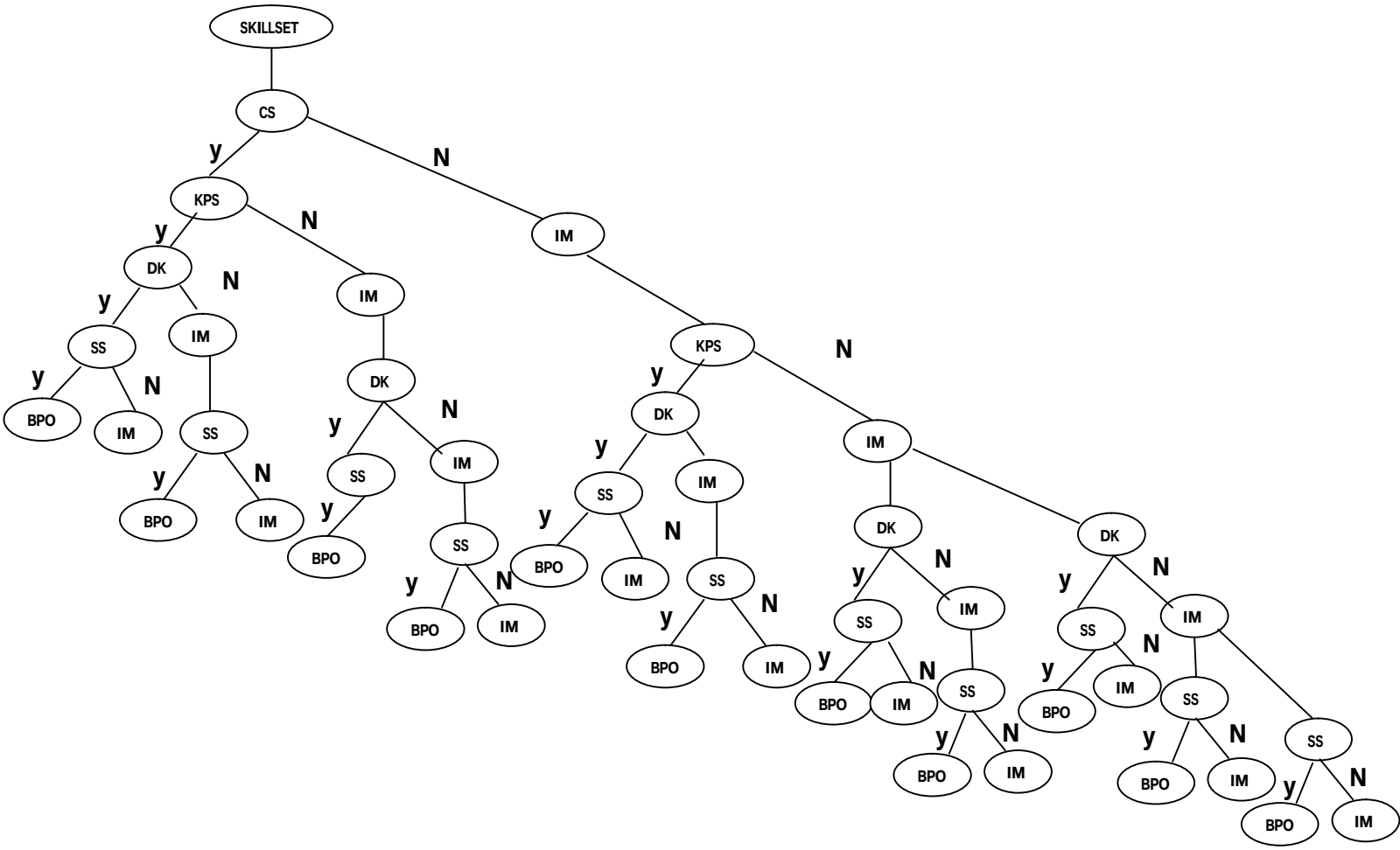


Figure 1. Decision Tree for the Skill Set in BPO -

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

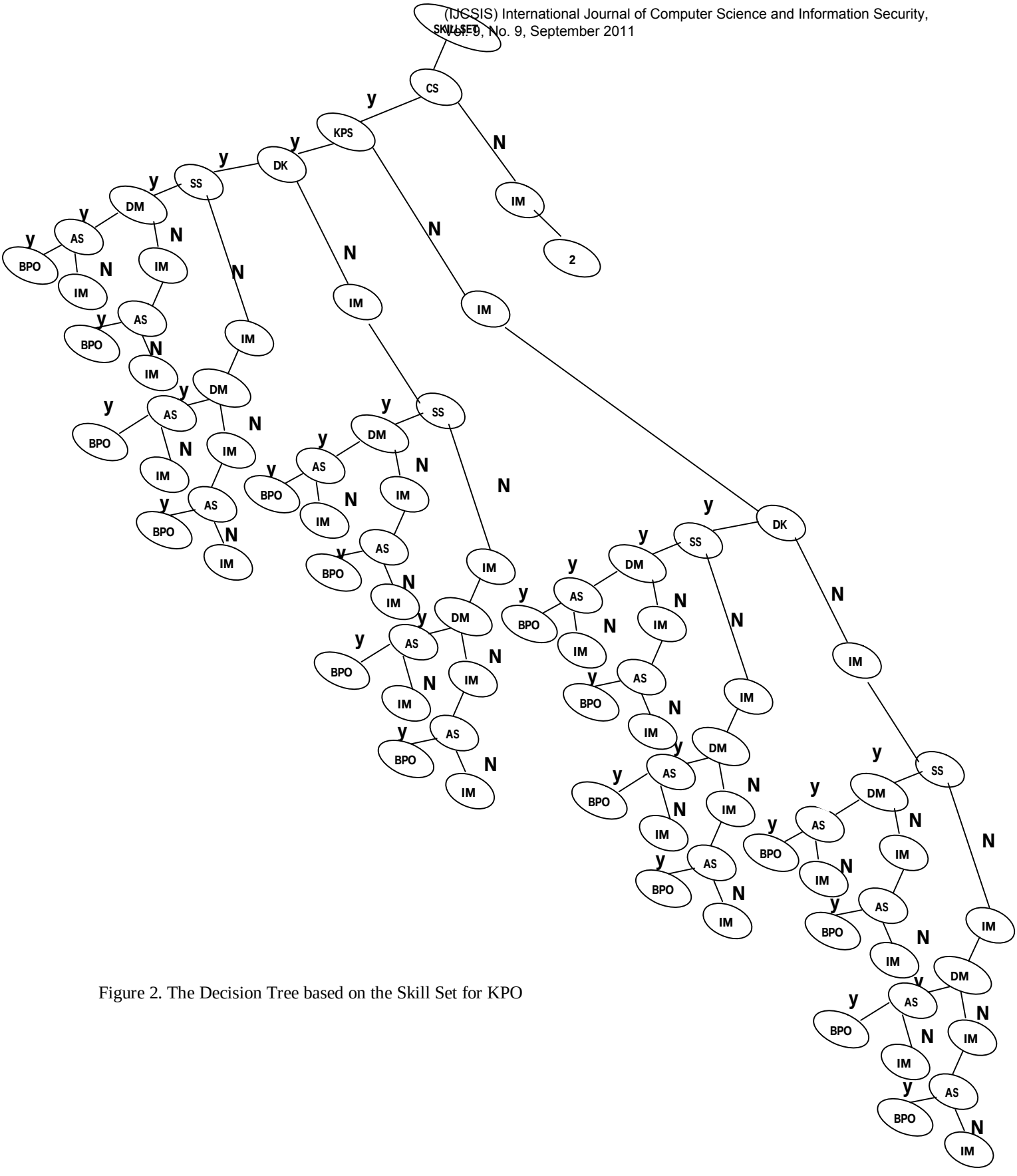


Figure 2. The Decision Tree based on the Skill Set for KPO

IV. CONCLUSION

This Paper aids in improvised decision making for the computer stream students in choosing their career path which paves the way to enter ITES industry. At present students lack in choosing the precise career path. In this paper ,a decision tree building model based on the various skill sets possessed by the students is presented . Firstly, the eligibility factor for BPO is evaluated by filtering the skill set(SOSS1).Secondly the eligibility factor for KPO is evaluated by filtering all the skill sets (SOSS2).The attributes are analyzed and the correct path is evaluated. With these factor decision tree is created. The results showed that, this method not only improves better decision making but also optimizes the structure of decision tree and gives provision for improving skill sets with alternative options. Due to the existence of various skill set possessed by students, how to choose the vital skill set and attribute is becoming a difficult task. In addition, the area of skill set with most reasonable attributes are worthy of exploration leading to various career path choice.

REFERENCES

[1]. Jiang Su and Harry Zhang "A Fast Decision Tree Learning Algorithm" American Associationa for Artificial Intelligence 2006.

[2] Sangjae Lee, "Using data envelopment analysis and decision trees for efficiency analysis andrecommendation of B2C controls" -Decision Support Systems 49 (2010) 486–497, ScienceDirect

[3] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, Sanfrancisco 2006

[4]Salsabil Trabelsi, Zied Elouedi *, Khaled Mellouli, "Pruning belief decision tree methods in averaging and conjunctive approaches" International Journal of Approximate Reasoning
46 (2007) 568–595, ScienceDirect

[5] Ilyes Jenhani *, Nahla Ben Amor, Zied Elouedi" Decision trees as possibilistic classifiers", International Journal of Approximate Reasoning
48 (2008) 784–807, ScienceDirect

[6].Han J,Kamber S. Data mining : Conepts and techniques .Morgan kaufman publishers 2006

[7]. D.Y. Sha . C-H .Liu, "Using Data Mining for Due Date Assignment in a a Dynamic Job Shop Environment", Int J Adv Manuf Technol(2005) 25:1164-1174

[8]. J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, Sanfrancisco 2001.

AUTHORS PROFILE

T.Hemalatha, is a research scholar in Bharathiar university,Coimbatore , India..She has published papers in international journals.Her area of interests are Decision Support System, Computer Application and Education Technology.

Dr. Ananthi Sheshasaayee received her Ph.D in Computer Science from Madras University,India. At present she is working as Associate professor and Head, Department of computer science, Quaid-e-Millath Government College for Women, Chennai.She has published 16 National and International journals. Her area of interest involve the fields of Computer Applications and Educational technology

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Real Time Transmission Protocol (RTP) and Real Time Transmission Control Protocol (RTCP) Library Implementation

Mohammad Monirul Islam
Computer Science and Engineering Department
Daffodil International University 102, Shukrabad, Dhaka-1207, Bangladesh

ABSTRACT— The system proposed is to create a library that will cover all the specifications given in RFC 3550.

I. INTRODUCTION

RFC 3550 is Transport Protocol for Real-Time Applications. RFC 3550 is RTP Version 2 protocol definition **[1]** proposed library would be used to develop applications that deal with transmission of real time audio and video data. RFC (**R**equest **F**or **C**omments) a document that describes the specifications for a recommended technology. Any RFC has specifications that are classified under following categories 'must', 'must not', 'required', 'shall', 'shall not', 'should', 'should not', 'recommended', 'may', 'may not' and 'optional'. Compliance with of an implementation with an RFC is measured by checking against req. Marked with above keywords. I intend to develop a library that would be compliant to the maximum possible extent. **[2]**

The Library will provide simple API's for user interface. Library will take care of Session Management. Library will handle error correction and data recovery for incoming and outgoing packets. The Library will be easily scalable. Library will be able to handle packet validations. The Library should provide error-handling mechanism.

II. Existing System and Need for System

Many implementations of Real Time Transmission Protocol (RTP) or Real Time Transmission Control Protocol (RTCP) are available in the market, but some of them are too specific to certain kind of applications and some are not easy to customize as per need of the application. One such RTP/RTCP library available under open GPL, The GNU General Public License (GNU GPL or simply GPL) is a widely used free software license, originally written by Richard Stallman for the GNU project. The latest version of the license, version 2, was released in 1991. The GNU Lesser General Public License (LGPL) is a modified version of the GPL, intended for some software libraries [3] which do not provide any kind of support. Any customization needs direct changes in library code, which requires complete understanding of library code. This library is not implemented using object oriented concepts. [4] To cope with these drawbacks there is a need of library, which remains private & easy to customize. This new implementation will be based on object oriented scenarios. Real-Time Transfer Protocol (RTP)

RTP was developed by the Audio/Video Transport working group of the IETF and has since been adopted by the ITU as part of its H.323 series of recommendations, and by various other standards organizations. The first version of RTP was completed in January 1996. RTP needs to be profiled for particular uses before it is complete; an initial profile was defined along with the RTP specification, and several more profiles are under development. Profiles are accompanied by several payload format specifications, describing the transport of a particular media format.

Real-Time Transfer Protocol consists of two major components:

1 Real Time Protocol (RTP): It carries real-time data.

2 Real Time Control Protocol (RTCP): It monitors the quality of service and conveys information about the participants **[2]**

III. The RTP Data Transfer Packet

RTP Sessions

A session consists of a group of participants who are communicating using RTP. A participant may be active in multiple RTP sessions—for instance, one session for exchanging audio data and another session for exchanging video data. For each participant, a network address and port pair to which data should be sent, and a port pair on which data is received identify the session. The send and receive ports may be the same. Each port pair comprises two adjacent ports: an even-numbered port for RTP data packets and the next higher (odd-numbered) port for RTCP control packets. The default port pair is 5004 and 5005 for UDP/IP, but many applications dynamically allocate ports during session setup and ignore the default. RTP sessions are designed to transport a single type of media; in a multimedia communication, each media type should be carried in a separate RTP session **[5]**

The RTP header has the following format:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

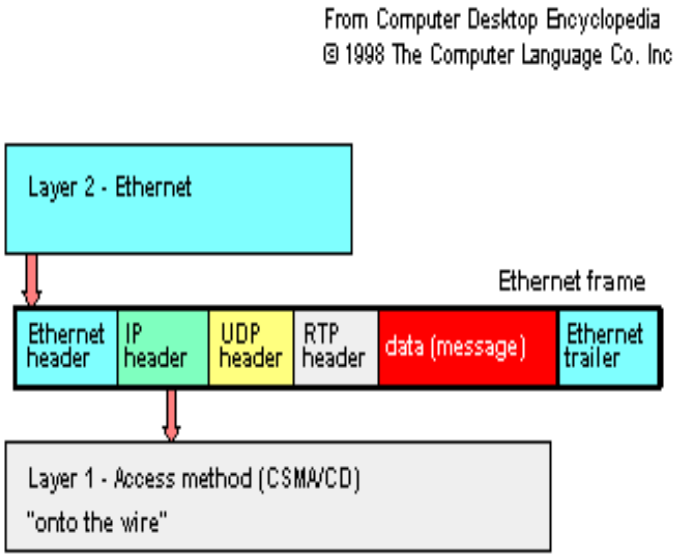


Fig: RTP header Format

Real Time Control Protocol (RTCP)

The RTP control protocol (RTCP) is based on the periodic transmission of control packets to all participants in the session, using the same distribution mechanism as the data packets. The underlying protocol must provide multiplexing of the data and control packets, for example using separate port numbers with UDP. RTCP performs four functions:

1. The primary function is to provide feedback on the quality of the data distribution. The feedback may be directly useful for control of data encodings that are currently used.

2. RTCP carries a persistent transport-level identifier for an RTP source called the canonical name or CNAME. Since the SSRC identifier may change if a conflict is discovered or a program is restarted, receivers require the CNAME to keep track of each participant. Receivers may also require the CNAME to associate multiple data streams from a given participant in a set of related RTP sessions.

3. By having each participant send its control packets to all the others, each can independently observe the number of participants. This number is used to calculate the rate at which the packets are sent.

4. A fourth, optional function is to convey minimal session control information, for example participant identification to be displayed in the user interface. This is most likely to be useful in "loosely controlled" sessions where participants enter and leave without membership control or parameter negotiation.

IV. RTCP Packet Format

This specification defines several RTCP packet types to carry a variety of control information:

SR: Sender report, for transmission and reception statistics from participants that are active senders

RR: Receiver report, for reception statistics from participants that are not active senders and in combination with SR for active senders reporting on more than 31 sources.

SDES: Source description items, including CNAME

BYE: Indicates end of participation

APP: Application-specific functions **[2]**

Objectives of System

- Library should provide simple set of functions (interface) that would be used by applications to transmit RTP/RTCP data.
- It should be easily portable across platforms.

Scope of Work

Library implementation will be according to the specifications defined in RFC3550.

RFC3550 defines certain rules that are mandatory for real time data transmission.

Operating Environment – Hardware and Software

V. Hardware Requirements

128 and above RAM

5GB or more Hard Disk..

Well Established Network environment.

VI. Software Requirements

Development ToolKit. C++ on Linux .

Operating System Linux 7.1 & above.

VII. LIBRARY IMPLEMENTATION

Library implements RTP/RTCP protocol by providing suitable APIs, each API functions as follows:

1. RTP_CreateSession()

This is the first API that application will call to establish a session. It takes parameters like user's self-IP address and pair of ports defined to receive RTP and RTCP

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

packets. Internally this API then creates participant list and adds self-entry in that list. Each participant entry in list contains information about participant and it's state. It also creates pair of UDP sockets through which user will send and receive RTP/RTCP packets. Finally it returns handle to session, which will be used as a parameter in subsequent APIs.

2. RTP_SetPayloadSize()

This API sets the payload size for RTP data packets. Payload size will be determined by payload format that is used by all session members. API internally allocates buffer of this size plus size of RTP packet header to carry RTP packet.

3. RTP_CalculateRTCPTimeInterval()

This API internally calculates RTCP time interval , the time interval is calculated by considering many session parameters like number of session members, number of active senders , number of receivers, bandwidth allocated to the session etc. On expiration of this interval participant may send RTCP packet.

4. RTP_SendRTPData()

Application will use this API to send media-data . API takes parameters as media-data and SSRC of participant to whom this data needs to be send. API internally creates RTP packet by feeling header fields and attaches this header to media-data. This packet then transferred to participant with given SSRC.After sending such a packet self-information in participant list will get updated.

5. RTP_RecvRTPData()

Application calls this API to receive RTP data. Internally API constructs the RTP packet from received raw data. Then the packet gets validated for various header fields. Once packet is valid only media-data and sender's SSRC is given to application. Then state of participant from which this packet is received is updated in participant list.

6. RTP_AddParticipant()

After receiving RTP/RTCP packet application can call this API to add the participant from which packet is received. API needs parameters such as CNAME, SSRC of participant, using such information internally participant list is checked to ensure that its entry is already there or not, if not then new entry gets created and initial values for that participant are set.

7. RTP_RemoveParticipant ()

Application can call this API when it wants to remove particular participant from the participant list. Internally the entry for that participant gets deleted from participant list.

8. RTP_SendRTCPReport ()

Application can use this API to send either Receiver Report or Sender Report packet of RTCP.API takes parameters as handle of current session, type of RTCP report and SSRC of destination participant. Internally this API will create appropriate report packet based on type specified. In this packet creation, RTCP common header will be filled, along with this it will also generate report blocks. Finally this packet will then get transferred to specified SSRC. Then self-information from participant list gets updated.

9. RTP_RecvRTCPReport()

Application can call this API to receive RTCP packet. Internally this API determines the incoming RTCP packets type and builds corresponding RTCP packet. Then information about participant from which this packet is received gets updated in participant list. Finally the structure that describes the received RTCP packet is given to the application.

10. RTP_SendRTCPByePacket()

Application calls this API when it wants to leave the session or when he finds the conflicting SSRC. API takes parameter as handle of current session and string that gives reason of saying BYE and length of the string.

11.RTP_CloseSession()

Application calls this API when it wants to close the current session by specifying the handle of session. API internally releases all the resources of session like participant list, session object etc.

CONCLUSION

The key standard for audio/video transport in IP networks is the Real-time Transport Protocol (RTP), along with its associated profiles and payload formats. RTP aims to provide services useful for the transport of real-time media, such as audio and video, over IP networks. These services include timing recovery, loss detection and correction, payload and source identification, reception quality feedback, media synchronization, and membership management.

By making use of suitable error detection and correction method, it is possible to transfer real time data on IP network using RTP protocol.

REFERENCES

[1] RFC 3551- http://rfc.net (Accessed on 20/08/06)

[2] RFC 3550, RFC 3551

[3] Wikipedia, the free encyclopedia - http://.wikipedia.org (Accessed on 10/08/06)

[4] www.gnu.org (Accessed on 14/08/06)

[5] Audio and Video for the Internet, -By Colin Perkins

[6] **Mohammad Monirul Islam**, is post graduated from London Metropolitan University, London, UK. At present he is working as Lecturer in Daffodil International University, Dhaka, Bangladesh. His major area of interset is :Programming languages, Databases, Advance database, Data mining and data warehousing, Real time data transmission, Internet application.

Towards Discriminant Analysis Modeling of Web 3.0 Design and Development for Students, Faculty and IT Professionals

S.Padma and Dr.Ananthi Seshasaayee

S..Padma,
Research scholar,
Bharathiar University
Coimbatore.
Assistant Professor
Vels University
Chennai .
INDIA.
padmanivasan@gmail.com

Dr.Ananthi Seshasaayee
Associate Prof. &Head,
Quaid-e-Millath Govt.
College for women,
Chennai
INDIA
ananthiseshu@gmail.com

Abstract

Web 3.0 is an evolving extension of the web 2.0 scenario. The perceptions regarding web 3.0 is different from person to person . Web 3.0 Architecture supports ubiquitous connectivity, network computing, open identity, intelligent web, distributed databases and intelligent applications . Some of the technologies which lead to the design and development of web 3.0 applications are Artificial intelligence, Automated reasoning, Cognitive architecture, Semantic web . An attempt is made to capture the requirements of Students, Faculties and IT professionals regarding Web 3.0 applications so as to bridge the gap between the design and development of web 3.0 applications and requirements among Students, Faculties and IT professionals. Discriminant modeling of the requirements facilitate the identification of key areas in the design and development of software products for Students, Faculties and IT professionals in Web 3.0.

Keywords : Web 3.0, Discriminant analysis , Design and Development ,Model

I INTRODUCTION

Web 3.0 is an extension of www, in which the information can be shared and interpreted by other software agent to find and integrate applications to different domains. Web 3.0 provides integrated real time application environment to the user. The applications are majorly involved in searching using semantic web, 3D web and are media centric. Web 3.0 supports pervasive components. Each component and its relations are represented below.

In web 3.0, web is transformed into database or Data Web wherein the data which are published in the web is reusable and can be queried. This enables a new level of data integration and application interoperability between platforms. It also makes the data openly accessible from anywhere and linkable as web pages do with hyperlinks. Data web phase is to make available structured data using RDF[1]. The scope of both structured and unstructured content would be covered in the full semantic web stage. Attempts will be to make it widely available in RDF and OWL semantic formats.

The driving force for web 3.0 will be artificial intelligence. Web 3.0 will be intelligent systems or will depend on emergence of intelligence in a more organic fashion and how people will cope with it. It will make applications perform logical reasoning operations through using sets of rules expressing logical relationships between concepts and data on the web. With the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

realization of the semantic web and its concepts web 3.0 will move into Service Oriented Architecture.

The evolution of 3D technology is also being connected to web 3.0 as web 3.0 may be used on massive scale due to its characteristics.

Web 3.0 is media centric where users can locate the searched media in similar graphics and sound of other media formats.

The pervasive nature of web 3.0 makes the users of web in wide range of area be reached not only in computers and cell phones but also through clothing, appliances, and automobiles.

II REVIEW OF LITERATURE

Claudio Baccigalupo and Enric Plaza discussed in the paper poolcasting : a social web radio architecture for Group Customization about Pool casting a social web radio architecture in which groups of listeners influence in real time the music played on each channel. Pool casting users contribute to the radio with songs they own, create radio channels and evaluate the proposed music, while an automatic intelligent technique schedules each channel with a group customized sequence of musically associated songs[2] . M.T.Carrasco Benitez discussed in the paper Open architecture for multilingual social networking about an open architecture for all the multilingual aspects of social networking. This architecture should be comprehensive and address well-trodden fields such as localization, and more advanced multilingual techniquesto facilitate the communication among users[3] . Autona Gerber, Alta van der Merwe, and Andries Barnard discussed in the paper A functional Semantic web architecture about the CFL architecture which depicts a simplification of the original architecture versions proposed by Bernes-Lee as a result of the abstraction of required functionality of language layers. Gerber argues that an abstracted layered architecture for the semantic web with well defined functionalities will assist with the resolution of several of the current semantic web research debates such as the layering of language technologies [4]. Ferda Tartanoglu val'erie Issarny, Alexander Romanovsky and Nicole Levy discussed in the paper Dependability in the web services architecture which lists about how to build dependable systems based on the web services architecture. It surveys base fault tolerance mechanisms and shows how they are adapted to deal with the specifics of the web in the light of ongoing work in the area[5]. Barry Norton, Sam Chapman and Fabio Ciravegna discussed in the paper developing a Service-Oriented Architecture to Harvest information for the Semantic web which discusses about the Armadillo architecture, how it is reinterpreted as workow templates that compose semantic web services and show how the porting of Armadillo to new domains, and the application of new tools, has been simplified[6].

III PROBLEM DEFINITION

The Design and Development of web 3.0 products are on the course. Due to the existence of the ambiguity in the requirements of Students, Faculty and IT professionals for structuring the web 3.0 products , bridging the gap between web 3.0 developers and Students, Faculty and IT professionals is required. The key factors for each of these three categories students ,faculty and it professionals are to be identified and their preference order is to be extracted.

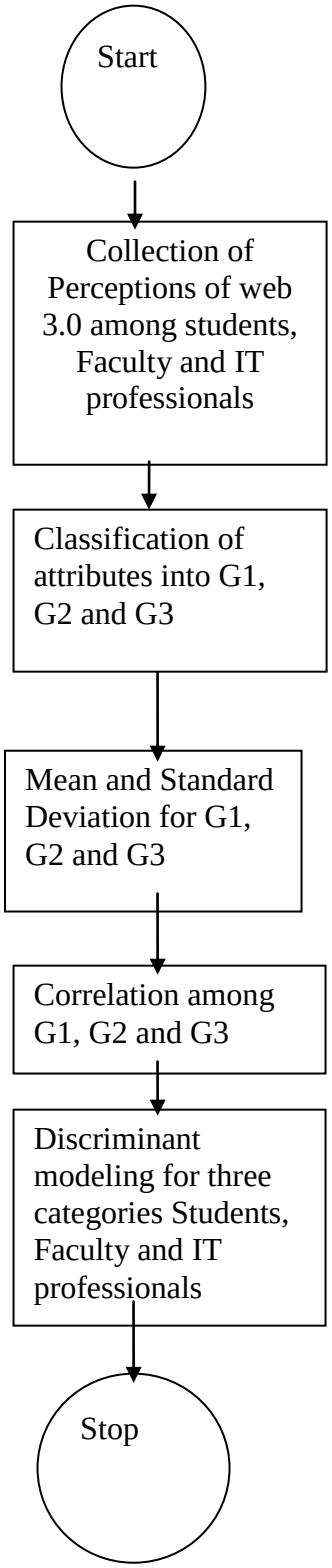
Let G1, G2, G3 denote the three groups in web 3.0 . The problem is to find the order of preferences of the three groups for the three categories Students, Faculty and IT professionals based on the attributes v1 , v2 , ….. vn included in these three groups G1, G2 and G3 to facilitate the design and development of applications in web 3.0 for the categories.

IV MATERIALS AND METHODS

We collected the perceptions of students ,Faculties and IT professionals inline with web 3.0 attributes. A five point scale was adapted which ranges from very low satisfaction , low satisfaction, Medium satisfaction, high satisfaction to very high satisfaction.

a. Block diagram of Web 3.0 discriminant modeling

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



b. Steps in Web 3.0 Discriminant modeling

a. Start
b. Collect the perceptions regarding the attributes of web 3.0 among the three categories Students, Faculty and IT professionals
c. Classification of attributes into three groups G1, G2 and G3.
d. Compute Mean and Standard Deviation for G1, G2 and G3
e. Correlation Coefficient among the groups G1, G2 and G3
f. Discriminant Modeling for the three categories Students, Faculty and IT professional
g. stop

c. Preprocessing

The data collected are verified for completeness. The missing values are replaced with the mean value.

d. Classification

The data collected from the three categories Students , Faculty and IT professionals based on the attributes 2D, 3D, Audio, Custom mash up, E decisions, Multilingual, Result as Mash up, Semantic Maps, Semantic Wiki, Software Agents, Speech recognition. Based on the functionality of the attributes they are grouped into G1 , G2 and G3. G1 comprises of Multilingual, Semantic maps, Edecisions, Semantic wiki and Software agents . G1 is termed as Applications . G2 comprises of 3D, Audio, 2D and Speech recognition. G2 is termed as Media. G3 comprises of Custom Mash up, Result as Mash up . G3 is termed as Output.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

e. Mean and Standard Deviation

TABLE 1. COMPARISION OF MEAN FOR THE THREE CATEGORIES

	MEAN		
CATEGORY	G1	G2	G3
STUDENTS	3.94049	3.5794	2.92
FACULTY	3.17	2.95	2.49
ITPROFESSIONALS	3.97	3.9	3.36

For all the three categories G1 Applications has higher mean when compared to all others.

TABLE II . COMPARISION OF STANDARD DEVIATION FOR THE THREE CATEGORIES

	STANDARD DEVIATION		
CATEGORY	G1	G2	G3
STUDENTS	0.57	0.52	0.53
FACULTY	0.57	0.49	0.45
ITPROFESSIONALS	0.58	0.55	0.39

The standard deviation for G3 are comparatively lower for faculty and IT professionals . Faculty and IT professionals have similar opinions about G3 – output. There is no significant difference in the standard deviation of students.

f. Finding the Correlation Coefficient

Correlation Coefficient reveals the nature relationship between the attributes.

The correlation coefficient for all pairs among the Groups are calculated using the following formula.[7]

Correlation(r) =[NΣXY - (ΣX)(ΣY) /
Sqrt([NΣX2 - (ΣX)2][NΣY2 - (ΣY)2])]
where

N = Number of values or elements
X = perception weightage for 1st group
Y = perception weightage for 2nd group
ΣXY = Sum of the product of first and Second group perceptions
ΣX = Sum of 1st group
ΣY = Sum of 2nd group
ΣX2 = Sum of square 1st group
ΣY2 = Sum of square 2nd group

TABLE III . CORRELATION AMONG GROUPS FOR STUDENTS

	STUDENTS		
	G1	G2	G3
G1	1	0.56	0.74
G2	0.56	1	0.51
G3	0.74	0.51	1

TABLE IV .CORRELATION AMONG GROUPS FOR FACULTY

	FACULTY		
	G1	G2	G3
G1	1	0.34	0.34
G2	0.34	1	0.28
G3	0.34	0.28	1

TABLE V. CORRELATION AMONG GROUPS FOR IT PROFESSIONALS

	ITPROFESSIONALS		
	G1	G2	G3
G1	1	0.3	0.42
G2	0.3	1	0.32
G3	0.42	0.32	1

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

It is evident that all the three groups G1, G2 and G3 are positively correlated to each other in all the three categories Students, Faculty and IT professionals.

g. Discriminant modeling on groups:

Discriminant Function analysis is a two step process.

Step 1: A set of discriminant functions are tested for its significance.

1.a. A Matrix of total variances and Co variances are constructed

1.b. A matrix of pooled within group variances and Co variances are constructed.

1.c. F test is performed on the two matrices constructed.

1.d Variable which have significantly different means across the groups are identified.

Step 2. Classification.

In the classification step classification of variables are done. DA automatically determines some optimal combination of variables so that the first function provides the most overall discrimination between groups and the second provides the second most and so on. The functions are independent or orthogonal [8].

V RESULTS AND DISCUSSION

TABLE VI . CANONICAL FUNCTION COEFFICIENTS

	Function	
	1	**2**
Applications (x1)	**3.247**	**-0.01**
Media (x2)	**0.885**	**0.408**
Output (x3)	**-0.547**	**1.887**
(Constant)	**-2.339**	**-7.452**

Based on the Canonical Discriminant Function coefficient, the linear discrminant equation can be written as

TABLE VII CLASSIFICATION RESULTS

$$Y = -2.339 + 3.247\, x_1 + 0.885\, x_2 - 0.547\, x_3 \quad (1)$$

$$Y = -7.452 - 1.017 x_1 + 0.408 x_2 + 1.887 x_3 \quad (2)$$

Based on (1) the following are the classification results.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Classification Results						
		Group	Predicted Group Membership			
			Students	Faculty	IT Professionals	Total
Original	Count	Students	218	7	8	233
		Faculty	4	347	6	357
		IT Professionals	3	12	421	436
	%	Students	93.5	7.3	4.7	100
		Faculty	1.1	97.2	1.7	100
		IT Professionals	0.7	2.8	96.6	100

a. 95.8% of original grouped cases correctly classified.

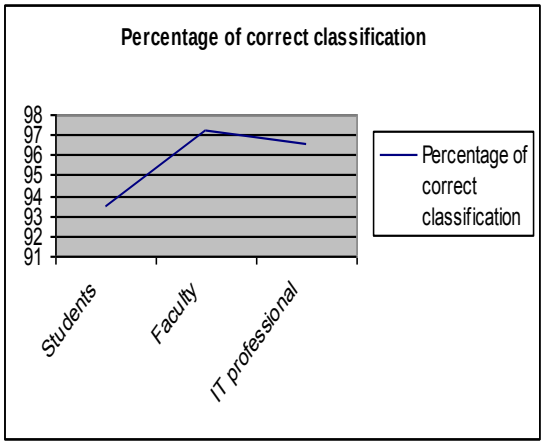


TABLE VIII CLASSIFICATION FUNCTION COEFFICIENTS

	Categories		
	Students	Faculty	ITProfessionals
Applications	14.048	3.743	4.818
Media	9.374	6.097	7.488
Output	8.074	7.66	12.521
(Constant)	-46.475	-29.192	-49.519

The order of preferences for the three categories is given below based on the above Classification Function Coefficients.

Order of Preferences among groups for Students

STUDENTS	APPLICATIONS	Multilingual (P1)
		Semantic map(P2)
		Semantic Wiki (P3)
		Edecisions (P4)
		Software agents (P5)
	MEDIA	
		2D (P6)
		3D (P7)
		Speech recognition (P8)
		Audio (P9)
	OUTPUT	
		Custom mash up (P10)
		Result as mash up (P11)

:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

From the above three tables the design and development of web 3.0 products specifically related to Students, Faculty and IT professionals can ensue the group preference orders and attributes . The products can be designed with the maximum attributes in the first group preference followed by lesser attributes in the second and third group.

VI CONCLUSION

The perceptions inline with web 3.0 are collected from students, Faculty and IT professionals. The data's are preprocessed , classified, Mean, Standard deviation and correlation coefficient are computed to understand the descriptive and Discriminant modeled. At the outset of evolving growth in Web 3.0 this model is an initiative for the of web 3.0 product design for Students , Faculty and IT professionals.

Order of preferences among groups for Faculty

FACULTY	OUTPUT	Custom mash up (P10)
		Result as mash up (P11)
	MEDIA	2D (P6)
		3D (P7)
		Speech recognition (P8)
		Audio (P9)
	APPLICATIONS	Multilingual (P1)
		Semantic map(P2)
		Semantic Wiki (P3)
		Edecisions (P4)
		Software agents (P5)

Order of preferences among groups for IT Professionals

ITPROFESSIONALS	OUTPUT	Custom mash up (P10)
		Result as mash up (P11)
	MEDIA	2D (P6)
		3D (P7)
		Speech recognition (P8)
		Audio (P9)
	APPLICATIONS	Multilingual (P1)
		Semantic map(P2)
		Semantic Wiki (P3)
		Edecisions (P4)
		Software agents (P5)

REFERENCES

[1]. en.wikipedia.org/wiki/Resource_Description_Framework

[2] Baccigalupo C.; Plaza E.; Poolcasting: a Social Web Radio Architecture for Group Customisation.Computer.org 2007.http://portal.acm.org/citation.cfm?id=1332471.1332755.

[3].]. Carrasco Benitez M.T;, Open architecture for multilingual social networking. Position paper for the W3C workshop on the Future of Social Networking [FSN], 15-16 January 2009.

[4]. . Aurona Gerber; Alta van der Merwe ; Andries Barnard; A functional semantic architecture.

http://www.eswc2008.org/final-pdfs -for-web-site/fisr-1.pdf.

[5]. Ferda Tartanoglu val'erie Issarny; Alexander Romanovsky and Nicole Levy er; Dependability in the web services architecture. 2007 .http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.771&rep=rep1&type=pdf.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[6]. Barry Norton; Sam Chapman ; Fabio Ciravegna; Developing a Service- Oriented Architecture to Harvest information for the Semantic web .2005.

[7]. http://easycalculation.com/statistics/learn-correlation.php

[8]. http://www.statpower.net/Content/312/Lecture%20Slides/Discriminant%20Analysis.pdf

[9]. Production of Cross Media Content for Multi-Channel Distribution, 2007, AXMEDIS apos;07 Third International Conference Volume ,Issue , 28-30 Nov . 2007 Page(s):115-122.

[10] Mantovani.F; VR Learning : Potential and Challenges for the Use of 3D Environments in Education and Traning . 2001.

[11]. Andr.P.Freire; Renata; P.M.Fortes ;An Evaluation of Web Accessibility Metrics based on their Attributes .2008.

[12]. Nuno Laranjeiro; Towards Fault Tolerance in Web Services Compositions .2007.

[13]. Dhiraj Joshi ; Ritendra Datta ; Ziming Zhuang ;PARAgrab : A Comprehensive Architecture for Web Image .2006.

[14]. Antonio Tapiador; Antonio Fumero, Joaqu'm Salvach'ua; Sandra Aguirre; A Web Collaboration Architecture .2006.

S.Padma , is a research scholar in Bharathiar university , Coimbatore. She has published 2 international journals . Her area of interests are Web mining.

Dr. Ananthi Seshasaayee received her Ph.D in Computer Science from Madras University. At present she is working as Associate professor and Head, Department of computer science, Quaid-e-Millath Government College for Women, chennai. She has published 17 international journals. Her area of interest involve the fields of Computer Applications and Educational technology.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Identifying Harmonics by Empirical Mode Decomposition for Effective Control of Active Filters During Electric Vehicle Charging

B. V. Dhananjay and T. Senthil

B.V. Dhananjay,
Research Scholar,
Vinayaka Mission's University,
Salem,India

Dr.T.Senthilkumar,
Professor, Automobile Engineering,
Bharathidhasan University,
Trichirapalli, India

***Abstract*-This paper provides Hilbert Huang Transform(HHT) method (an empirical mode decomposition(EMD)) for identifying the presence of harmonics during electric vehicle battery charging when harmonics are generated into the electric line, due to switching actions of the power electronics. Activation of the active filters based on the difference between load current and fundamental current measured from the line is done. By using active power filter (APF) injection of the required current to minimize the harmonics is done. As part of simulation, the accuracy of the HHT is above 95%. By correctly recognizing the harmonics using HHT and injecting the compensating current into the line, the charging time of the battery can be reduced. The reduction in the charging time also depends on the battery condition.**

Keywords-Hilbert Huang Transform; active power filter.

I INTRODUCTION

The battery is the primary source of electrical energy. It stores chemicals. Two different types of lead in an acid mixture react to produce an electrical pressure. This electrochemical reaction changes chemical energy into electrical energy. A battery can be of primary cell, secondary cell, wet charged, dry charged and low maintenance type. A fully charged battery contains a negative plate of sponge lead(Pb), a positive plate of lead dioxide(Pbo_2) and an electrolyte of sulphuric acid (H_2So_4) and water (H_2o). During charging, sulphate leaves the plates and combines with hydrogen(H_2) to become sulphuric acid (H_2So_4). Free oxygen combines with lead on the positive plate to form lead dioxide. Gassing occurs as the battery nears full charge and hydrogen bubbles out at the negative plates, oxygen at the positive. Factors affecting charging are temperature, state of charge, plate area, impurities, gassing.

Electric vehicles (EV) will become an attractive alternative to internal combustion engine vehicles in the event that their range can be extended. One way to achieve this in the short term is to provide a fast charger infrastructure. Such a structure would provide greater mobility for the EV user, since during short stops (<1 hour) the EV batteries could be charged from typically 20 to 80 % of nominal charge. This would significantly extend the EV range.Fast charger infrastructure cost is high. Chargers adversely affect the grid power quality due to presence of power electronic loads like diode rectifiers and thyristor bridge converters in the distribution network that result in voltage distortion and current harmonics, (Akagi 1996).

High increase of problems in the electric power distribution networks due to the presence of harmonics. Loads that use switching control with semiconductor devices are the main cause. One of the most important tools for correcting the lack of electric power quality are the active power filters (APF), (Udom et al. 2008). The objective of this work has been proving that back propagation neural networks, previously trained with a certain number of distorted waveforms, are an alternative to the rest of the techniques used and proposed at the present time for controlling the APF's, as the ones based on the use of the Fast Fourier Transform (FFT). A large number of these control techniques are based on ANN's, (Pecharanin et al. 1994).

II MATERIALS AND METHODS

A *Materials*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 1shows a three-phase diagram of an HHT controlled shunt APF. A load current signal $i_L$is acquiredand used by the ANN to obtain the distortion current waveform as reference signal for the control of the APF.The power converter injects the necessary compensation current $i_L$in the power circuit, achieving thus a sinusoidal source current.

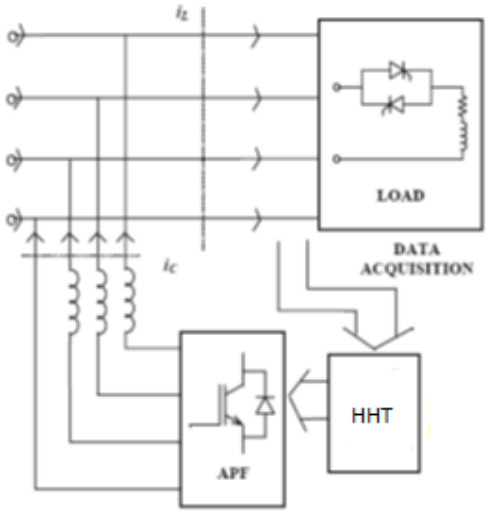


Figure 1 APF control using HHT

B *Methods*

Empirical Mode Decomposition (Huang) and Hilbert Transform

A signal can be analyzed in details for its frequency, amplitude and phase contents by using EMD followed by HT (Jayasree et al. 2010 and Stuti et al. 2009), The EMD produces the mono components called IMFs from the original signal. In a given frame of signal, there can be many IMFs. Each IMF will contain a wave form of different amplitude. Hilbert Transform is applied on an IMF to obtain, IF and IA. It is mandatory that a signal be symmetric regarding the local zero mean, and should contain same number of extreme and zero crossings.

The steps involved in EMD of a signal X(t) with harmonics into a set of IMFs are as follows.

1. Identify all local maxima of X(t). Connect the points using a cubic spline. The interpolated curve obtained. The upper line is called the upper envelope (Maximum_envelope).
2. Identify all local minima of X(t) connect the point using a cubic spline.. The lower line is called the lower envelope (Minimum_envelope) obtained by cubic spline.
3. Compute the average by:

$$M = \frac{(a+b)}{2} \quad (1)$$

Where a = Maximum_envelope and b = Minimum_envelope.

4. Obtain a new signal using the following equation:

$$h_{11}(t) = X(t) - M_{11}(t) \quad (2)$$

Where $h_{11}(t)$ is called first IMF. Subsequent IMF's had to be found if there are some overshoots and undershoots in the IMF. Hence, the envelope mean differs from the true local mean and $h_{11}(t)$ becomes asymmetric.

In order to find the additional IMF's, $h_{11}(t)$ is taken as the new signal. After n^{th} iteration, we have:

$$h_{1n}(t) = h_{1(n-1)}(t) - M_{1n}(t) \quad (3)$$

Where $M_{1n}(t)$ is the mean envelop after the n^{th} iteration and $h_{1(n-1)}(t)$ is the difference between the signal and the mean envelope at the $(k-1)^{th}$ iteration.

5. Calculate C2F as follows:

$$C2F_1 = IMF_n \quad (4)$$

Where IMF_n = final IMF obtained

$$C2F_2 = IMF_n + IMF_{(n-1)} \quad (5)$$

Similarly,

$$C2F_n = IMF_n + IMF_{(n-1)} + \ldots\ldots + IMF_1 \quad (6)$$

Where $C2F_n$ is the original signal.

6. Calculate F2C as follows:

$$F2C_1 = IMF_1 \quad (7)$$

$$F2C_2 = IMF_1 + IMF_2 \quad (8)$$

$$F2C_n = IMF_1 + IMF_2 + \ldots\ldots + IMF_n \quad (9)$$

Where $F2C_n$ is the original signal.

7. Hilbert transform is applied for each IMF and analytical signal is obtained. A complex signal is obtained from each IMF:

$$\text{Analytic(IMF)} = \text{real(IMF)} + \text{imag(IMF)} \quad (10)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

8. Instantaneous frequencies are obtained from analytical signal using

$$IF = \frac{0.5 \times (\text{angle}(-X(t+1) \times \text{conj}(X(t-1))) + \pi)}{2 \times \pi} \quad (11)$$

9. Instantaneous amplitudes are obtained from the analytical signal using the following

$$IA = \sqrt{\text{real}(IMF)^2 + \text{imag}(IMF)^2} \quad (12)$$

III EXPERIMENTAL SIMULATION

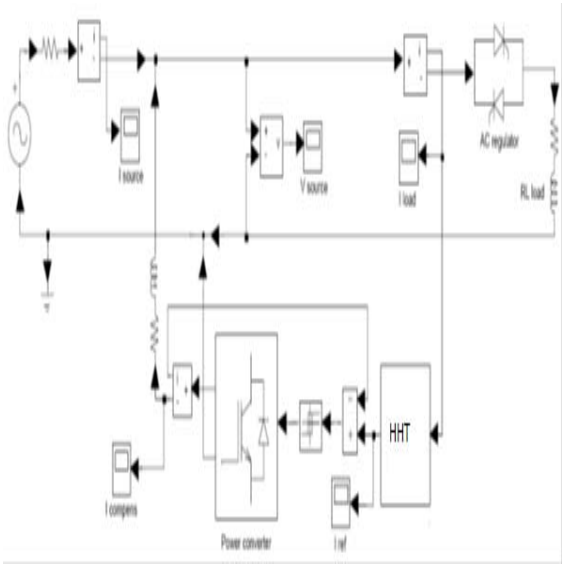


Figure 2 Power circuit with HHT controlling active power filter

The model of Figure2 has been created using Matlab 10. Different sets of parameters have been employed at the power circuit and APF. In most cases the reference current obtained by the HHT controller was accurate enough to enable the APF to compensate harmonic distortion. If an elevated content of high order harmonics were present in the load current, the HHT controller helps in obtaining reference signal.

System parameter used:
Power circuit: Phase voltage = 220 VRMS , Frequency = 50 Hz , Source resistance = 0.1 Ω, Load resistance = 20 Ω, Load inductance = 30 mH
APF: Vdc = 500 V, R = 10 Ω, L = 30 mH, Switching frequency = 40 KHz.

IV RESULTS AND DISCUSSION

More than 500 different harmonics waveforms (Figure 3) have been used in HHT analysis with different load changes.The results of a simulation are presented where a step load change occurs at time 60 ms. One additional resistance is connected in parallel with the load, increasing the total load current.

Figure 4 shows the EMD process. In the sample harmonics signal considered, only one instantaneous mode function is present. A flat residue signal is also presented. This plot is only for 1000 samples. This will be repeated for the remaining length of the signal. Figure 5 shows the extraction of different signals present from the fine level to coarse level. Similarly, Figure 5 shows the extraction of different signals present from the coarse level to fine level. Figure 7 presents the instantaneous frequency present in every sample of the signal. Figure 8 presents the instantaneous amplitude present in every sample of the signal. Figure 9 and Figure 10 presents statistical values of instantaneous frequencies and instantaneous amplitudes. Based on the statistical values, the amount of harmonics will be estimated and appropriately, required compensating current will be injected into the line.

V CONCLUSION

A Hilbert Huang Transform method has been used at the control of a shunt active power filter. Based on the amount of harmonics recognition, the APF is activated. By correctly injecting the compensating current into the line, the charging time of the battery can be reduced. The circuit has to be verified with the implementation of HHT in real time for improved charging of the EV battery. The reduction in the charging time also depends on the battery condition.

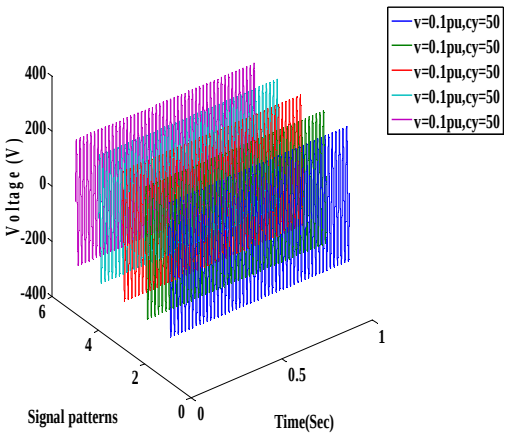


Figure 3 Sample plot of signals for harmonics

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

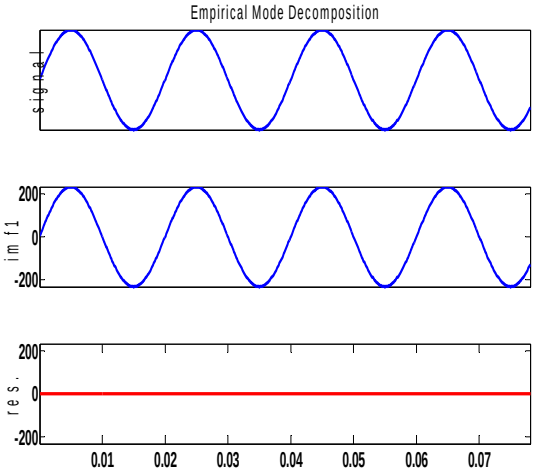


Figure 4 Empirical mode decomposition

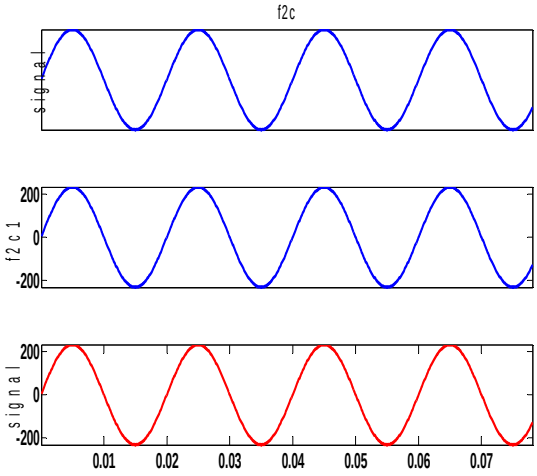


Figure 5 Fine to coarse signals

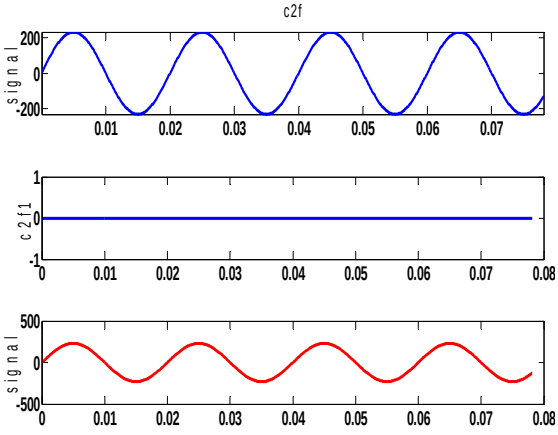


Figure 6 Coarse to fine signals

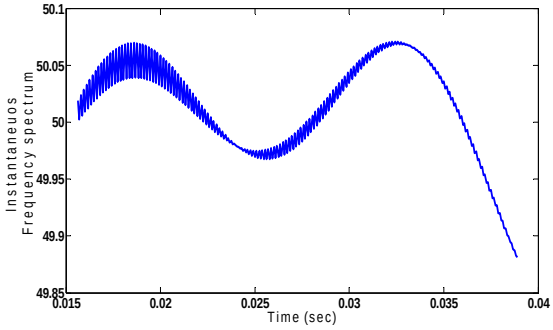


Figure 7 Instantaneous frequency

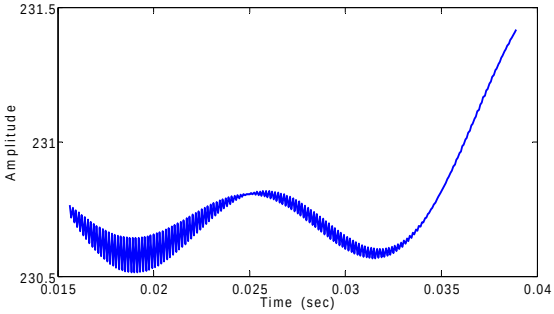


Figure 8 Instantaneous amplitude

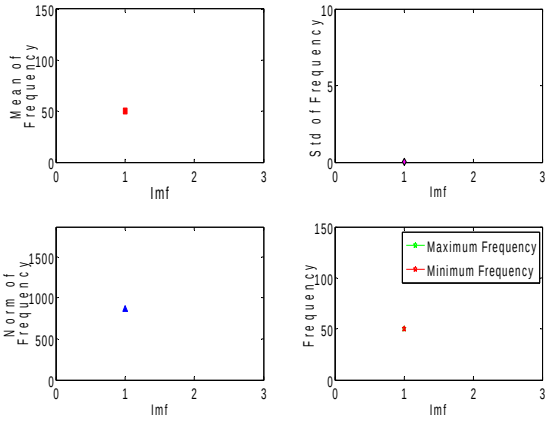


Figure 9 Statistical values of the instantaneous frequencies

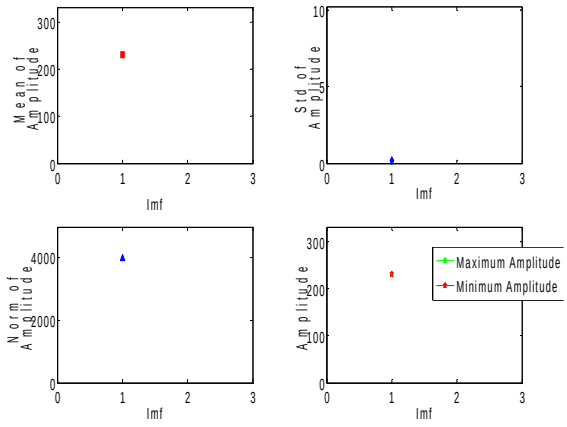


Figure 10 Statistical values of the instantaneous amplitudes

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] HAkagi., "New Trends in Active Filters for Power Conditioning", IEEE Trans. on Industry Applications, vol. 32, No 6, Dec. 1996, pp 1312-1322.

[2] TJayasree., DDevaraj., RSukanesh.,"Power quality disturbance classification using Hilbert transform and RBF networks", Neurocomputing , Vol 73 Issue 7-9, March, 2010, pp. 1451-1456.

[3] NPecharanin., MSone., HMitsui., "An application of neural network for harmonic detection in active filter",IEEE World Congress on Computational Intelligence.,IEEE International Conference on Neural Networks, Vol.6, pp. 3756-3760, 1994.

[4] StutiShukla, S. Mishra, and BhimSingh,"Empirical-Mode Decomposition With Hilbert Transform for Power-Quality Assessment", IEEE transactions on power delivery, Vol. 24, No. 4, October 2009.

[5] Udom. Khruathep, SuittichaiPremrudeepreechacharn, YuttanaKumsuwan,"Implementation of shunt active power filter using source voltage and source current detection", *IEEE*, pp.2364-2351, 2008.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Multilevel Representation of Repository for Software Reuse

C.V.Guru Rao
Professor & Head
Department of Computer Science and Engineering,
SR Engineering college, Warangal,
Andhra Pradesh, India – 506 371
guru_cv_rao@hotmail.com

P.Niranjan
Associate Professor,
Department of Computer Science and Engineering,
Kakatiya Institute of Technology and Science
Warangal, Andhra Pradesh, India – 506 015
npolala@yahoo.co.in

***Abstract*-- Effective software Reuse will be due to classification schemes used on software components that are stored into and retrieve from a software repository.**

This work proposes a new methodology for efficient classification and retrieval of multimedia software components based on user requirements by using attribute and faceted classification schemes. Whenever a user desires to trace a component with specified characteristics (Attributes) are identified and then compared with the characteristics of the existing components in repositories to retrieve relevant components. A web based software tool developed here to classify the multimedia software components is more efficient.

Keywords: Software Reuse, Classification Schemes, Reuse Repository.

I. INTRODUCTION

Software reuse is the use of engineering knowledge or artifacts from existing software components to build a new system [11]. There are many work products that can be reused, such as source code, designs, specifications, architectures and documentation. The most common reuse product is source code.

Software components provide a vehicle for planned and systematic reuse. Nowadays, the term component is used as a synonym for object most of the time, but it also stands for module or function. Recently the term component-based or component-oriented software development has become popular. Systematic software reuse influence the whole software engineering process. The ability to develop the new web based applications with in a short time is crucial to the software companies. For this reason it is vital to share and reuse the efficient programming experiences as well as knowledge in a productive manner.

A software component is a well-defined unit of software that has a published interface and can be used in conjunction with components to form larger unit [3].

To incorporate reusable components into systems, programmers must be able to find and understand them. If this process fails, then reuse cannot happen. Thus, to represent these components and index them is a challenge. Therefore to find them easily and understand the function are two important issues in creating a software tool for software reuse. Classifying software component allows reusers to organize collections of components into structures that they can search easily. Successful reuse requires proper classification and retrieval mechanisms to possess a wide variety of high quality components that are understandable.

Multimedia technology enables information to be stored in a variety of formats. Therefore very effective presentation of software components can be made. Understanding behavior of a component is very important for increasing the user's confidence before reuse the retrieved software component with different qualities from the library. Multimedia presentation will allow the users to better understand the software components.

Existing techniques are mainly focusing on representation issue of software components in software repositories. But they ignore the presentation of the software component semantics. In this paper an approach for integrated classification scheme with very effective presentation of reusable software components is presented. A software tool is developed to classify multimedia software components. Experimentally demonstrated the software tool is highly efficient.

The paper is organized as follows. Section 2 illustrates survey of related research work. The proposed classification technique to store and retrieve components is explained in section 3. Section 4 brings out the details of experimentation carried out on the proposed classification method. The experimental results are demonstrated in section 5. Section 6 concludes the work and followed by its references.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. RELATED RESEARCH

In the recent past research on software reuse has been focusing on several areas: examining programming language mechanisms to improve software reusability; developing software processes and management strategies that support the reuse of software; also, strategies for setting up libraries containing reusable code components, and classification and retrieval techniques to help a software professional to select the component from the software library that is appropriate for his or her purposes.

Earlier the research on software reuse was much focused on identifying reusable artifacts, storage and retrieval of software components. It had attracted more attention as it was essential for software developers.

A. Existing Software component Classification and Retrieval Techniques

"A classified collection is not useful if it does not provide the search-and-retrieval mechanism and use it" [10]. A wide range of solutions to the software component classification and retrieval were proposed and implemented. At different times, based on available software systems and also on researchers' criteria, software reuse classification and retrieval approaches are observed with minor variations.

Ostertag et al. [24] reported three approaches for classification. First is a free-text keywords next one is that a faceted index and the last one is semantic-net based. Free text based approach use information retrieval and indexing technology to automatically extract keywords from software documentation and index items with keywords. The free-text keyword approach is simple and an automatic process. But this approach curtails semantic information associated with keywords. Therefore it is not a precise approach. In faceted index approach, experts extract keywords from program descriptions and documentation. They arrange the keywords by facets into a classification scheme, which is used as a standard descriptor for software components. Mili et al [6] classifies search and retrieval approaches into four different types:

1) simple keyword and string match; 2) faceted classification and retrieva curtailsl; 3) signature matching; and 4) behavior matching. The last two approaches are cumbersome and inefficient.

Mili et al [6] designed a software library in which software components are described in a formal specification: a specification is represented by a pair(S, R), where S is a set of specification, and R is a relation on S.

The faceted classification scheme for software reuse proposed by Prieto-Diaz and Freeman [10] relies on facets which are extracted by experts to describe features about components. Features serve as component descriptors, such as the component functionality, how to run the component, and implementation details. To determine similarity between query and software components, a weighted conceptual graph is used to measure closeness by the conceptual distance among terms in a facet.

Girardi and Ibrahim's [25] solution for retrieving software artifacts is based on natural language processing. Both user queries and software component descriptions are expressed in natural language. Natural language processing at the lexical, syntactic and semantic levels is performed on software descriptions to automatically extract both verbal and nominal phrases to create a frame-based indexing unit for software components.

B. Factors Affecting Software Reuse Practices

Even though a substantial number of components are becoming common with repositories being developed, there are several problems with software reuse. First, a variety of components must be made available for reuse, which is maintained in a repository.

Next, the classification factors used to categorize the components play a vital role in the component reuse. Each component is annotated with a brief description of its role. Classification of components is done based upon pre-defined classifiers i.e. classification factors.

Further, the component vendors are making great strides in facilitating the distribution of components; no single vendor has emerged as the leader in providing a comprehensive solution to the search and retrieval problem. The size and organization of the component repositories further exacerbates the problem.

Finally, even if repositories are available, there are no easy or widely accepted means for searching for specific components to satisfy the users' requirements.

Software reuse deals with the ability to combine separate independent software components to form a larger unit of software.

Once the developer is satisfied with the component he had retrieved from library, then it is added to current project under development.

Literature reveals many methods for developing multimedia applications and processing multimedia data.

Various uses for multimedia annotation have been identified for computer based training and narration [5].

The aim of the good component retrieval system is to locate either the component required or

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the closest match in the shortest amount of time using a suitable query.

C. *Existing System Architecture*

Existing techniques use the architecture shown in the Figure 1. In this architecture classification and retrieval system relies upon single database interface to manage both storage and retrieval process. If number of components in the database are more, then searching method will become more inefficient.

In existing architectures software reusable components are directly stored in database. There is no special control and management of components. So retrieving of suitable components in a particular reuse scenario becomes tedious. This also facilitates to perform different operations like frequent component set and version control are becomes easy.

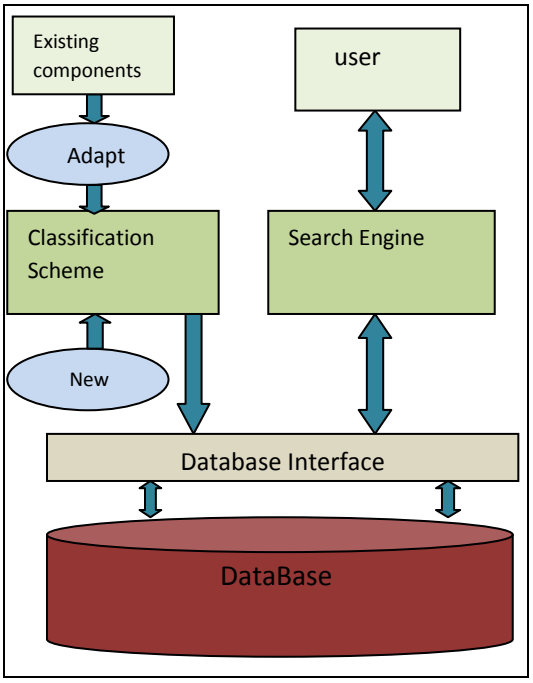


Figure 1. Existing System Architecture

III. PROPOSED SYSTEM

A. *Proposed Architecture*

Existing software components in the repository can be directly classified in the classification scheme into one among the above specified classifications presented in the previous section and then stored into a repository. Sometimes they need to be adapted according to the user requirements. As classification scheme inherently affect the classification efficiency due to the techniques in the previous section. New designs of software components for reuse are also subjected to classification scheme before storing them into a repository. User will retrieve his desired component with required attributes from the repository.

The existing architecture is inefficient when the number of components in the database are more.

To overcome this lacunae a modified architecture is proposed as shown in Figure 2. A dedicated repository is used to store and manage component details with multimedia information.

In the proposed architecture a separate reuse repository is responsible to control and manage all components. It ensures the quality of components and availability of necessary documentation and helps in retrieving suitable components with detailed description. This amounts to centralized production and management of software reusable components.

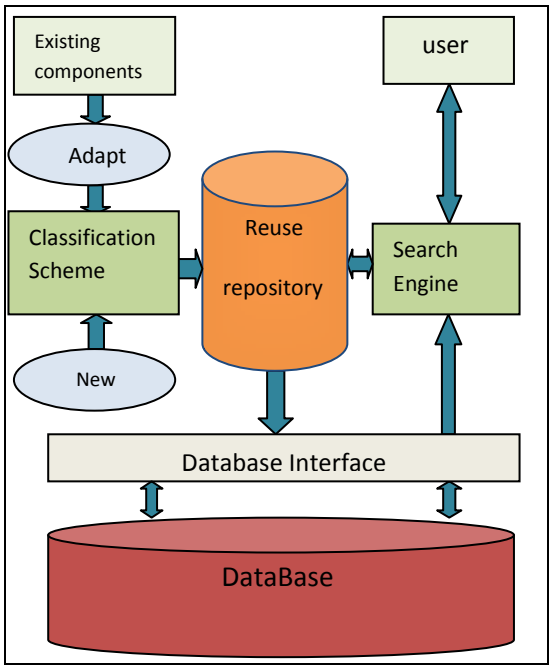


Figure 2. Proposed System Architecture

B. *Proposed Classification Scheme*

An Integrated Classification Scheme for Reusable Software Components with Multimedia Presentations is proposed. In this scheme an audio presentation is the combination of one or more classification techniques. It is likely to enhance the classification efficiency. This will give rise to the development of a software tool to classify a software component and build a reuse repository.

Integrated classification scheme which combines the faceted classification scheme to classify components with the following attribute values.

- Operating system
- Language, Function

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Inputs
- Outputs
- Domain
- Version

The software tool to be developed is aimed to provide an user friendly interface for browsing, retrieving and inserting components. Two separate algorithms for searching and another for inserting components are developed to work with this software tool.

Algorithm 1:

Component Insert(Component facet and attributes)
Purpose: This algorithm inserts a component into the reuse repository with integrated classification scheme attributes.
Component Insert:
Input:(component attributes, component)
Output: (success , failure)

Begin
Step1: Enter attribute values;
Step2: If (component attributes <>
existing components attributes)
Then
Store = success;
Else
Store = failure;

Step3: If (Store = success) Then
Component is successfully inserted into repository;
Else
Component already exists;
End.

The insert algorithm stores the newly designed or adapted existing component intoa repository. When component attributes are compared with existing component attributes in a repository. If component with this description is found then component is inserted successfully, otherwise component not inserted in repository and exits giving message that component already exists.

Algorithm 2:

Search_Component(Component facet and attributes)
Purpose: This algorithm searches for relevant components with given component facet and attributes from reuse repository.
Component Search:
Input: (component attributes)
Output: (relevant components)
Begin
Step1: Enter attribute values.
Step2: If (Any of the component attribute values = Repository components attributes) Then
Retrieves matching components from repository ;
Else
No matching components found;
End.

The search algorithm accepts facet of a component and attribute values from user intern it retrieves relevant components from repository.

C. Implementation

The above algorithms are implemented as the following modules and integrated as software tool..

a. User Interface

This module is designed to build a clearly defined, understandable documentation and with concise interface specifications. A graphic user interface is designed to select options like insert a component, delete a component and search for a component. Through this interface the user can easily submits his desired preferences for various operations.

b. Query Formation

The user preferences are captured to insert a component into repository or search for a component from a repository and a query is formed. Suppose a user desirous of searching a component may enter some keywords. He may also select some list of attributes from the interface. The query formation module should accept all the keywords entered and form the query using those keywords.

c. Query Execution

In this module user query will be executed and results are displayed. Suppose if user query is to retrieve components from repository then on query execution all the components which satisfy the criteria that is specified by user are displayed.

The results displayed give full details. Now the user can select his choice of component to download or save a component in the location specified by the user.

IV. EXPERIMENTATION

The software tool provides the options to store or retrieve components from repository. The following test cases are described when executed together with the algorithms explained in previous section

Sample test cases:

Case 1. Inserting a software component into reuserepository.

Component-id : 009
Operating system: Windows
Language , Function: Java , Sorting

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Input : Data items
Output : Sorted data items
Domain : Educational
Version : 2.0
Result: Insertion of a software component is successful.

In this test case, a given component attributes are captured and compared with components in the repository. The search algorithm does not find a matching component in the repository. Therefore, this component inserted into the repository and it results in successful insertion of component into repository.

Case 2. Inserting a component into reuse repository.

Component-id : 018
Operating system: Windows
Language , Function: Java , Sorting
Input : Data items
Output : Sorted data items
Domain : Educational
Version : 2.0

Result: This software Component is already exists in the reuse repository.

In this test case a given component attributes are captured and compared with components in the repository. The algorithm finds a matching component in the reuse repository. Therefore this software component is not inserted into the reuse repository. A message is displayed that the software component already exists in the reuse repository.

Case 3. Retrieving a software component from the reuse repository

Component-id : -
Operating system: -
Language , Function: Java , Sorting
Input : -
Output : -
Domain : -
Version : -

Result:

Comp-Id	version	
003	3.0	Download
018	2.0	Download
020	1.0	Download

In this test case language and function attributes are captured and compared with software components available in reuse repository. The algorithm found three relevant software components in the reuse repository. The results are displayed with full details of software components retrieved from reuse repository.

Case 4. Retrieving a software component from reuse repository.

Component-id : -
Operating system: Unix
Language , Function: Java , -
Input : -
Output : -
Domain : -
Version : -

Result: Full specifications of software component are not passed. Software component retrieval is failure.

In above test case total facet attributes are not given only language attribute is given. The search algorithm displays a message that function facet is not mentioned.

The experimental test cases are conducted with our integrated classification scheme algorithms and results are compared with existing schemes and result charts are presented in next section.

V. RESULTS

The performance is evaluated with different test results and compared with existing schemes.

Search effectiveness refers to how best a given method supports finding relevant items in given database. This may be number of relevant items retrieved over the total number of items retrieved. The following box-plots in Figure 3 illustrates the performance of search in existing classification

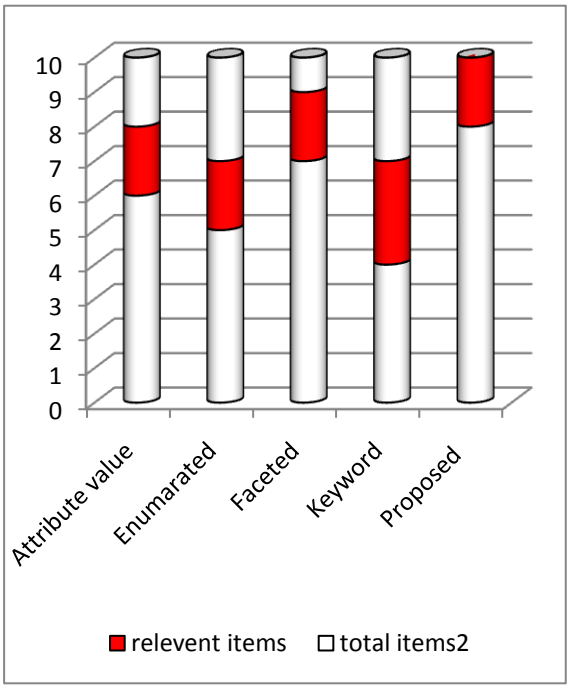


Figure 3 . Finding Relevant Components

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

schemes and integrated classification scheme on the horizontal axis for the number of data items as mentioned on the vertical axis. Total data items retrieved are shown with white color and colored area indicates the percentage of relevant items among all the retrieved data items.

Faceted classification scheme marked highest performance of search among all the existing classification schemes. Keyword classification scheme registered the lowest performance. Whereas our proposed integrated classification scheme out performed to retrieve more relevant items in comparison to all those existing schemes.

Search time is the length of time spent by a user to search for a software component. The following box-plots in Figure 4 gives search time consumed by the existing classification schemes and Integrated classification scheme.

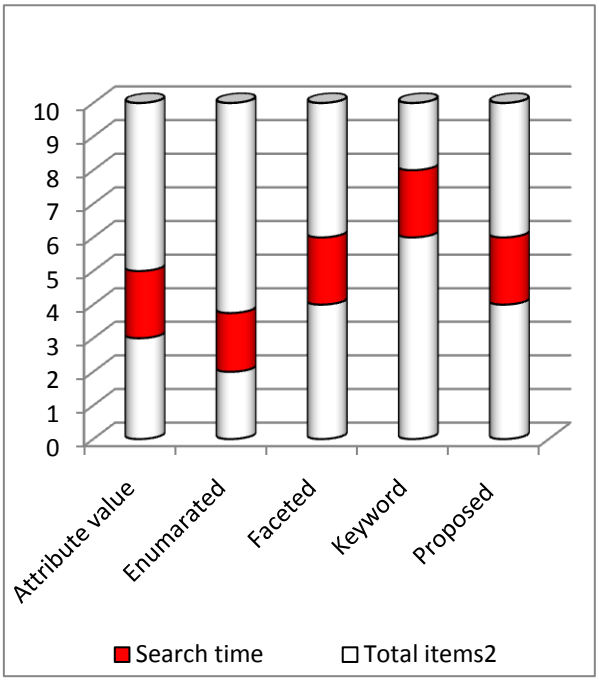


Figure 4. Search Time of Components

Existing classification schemes togher with proposed and Integrated classification scheme on the horizontal axis and the search time consumed in each method on the vertical axis. Total data items retrieved are shown with white color and colored area indicates the search time to retrieve those data items.

VI. CONCLUSION

This integrated classification scheme with multimedia presentation most efficient retrieval method over existing schemes. The relevant components for software reuse from the software repositories are presently drawn.. The solution realized here will suit to all the needs of various software developers in the industry.

The possibilities of further up gradation according to additional software requirements of the clients is not ruled out due to software reuse.

REFERENCES

[1] **S.** Arnold and **S.** Stepoway. The Reuse System: Cataloguing and Retrieval of Reusable Software. *Proceedings of COMPCONS'87,* 1987, **pp.** 376-379.

[2] T. Isakowitz and R. Kauflkan. Supporting Search for Reusable Software Components. *IEEE Transactions on Software Engineering.* Vol. 22, No 6, June 1996.

[3] B. Burton, R. Aragon, **S.** Bailey, K. Koehler and L.Mayes. *The Reusable Software Libra **y*** **.** IEEE Software.

[4] A. Al-Yasiri. *Domain Oriented Object Reuse based onGeneric Software Architectures,* Ph.D. Thesis, Liverpool John Moores University, May 1997.

[5] S. Bailin. *Applying Multimedia to the Reuse of Design Knowledge.* Paper available at: http://www.umcs.main.edu/-ftp/wisr/wisr8/papers/bailin/bailhhtml. 1997.

[6] R. Mili, **A.** Mili and R.T. Mittermeir. Storing and Retrieving Software Based Components: A Refinement Based System. *IEEE Transactions on Software Engineering.* Vol. 23, No 7, July 1997.

[7] Y. Maarek. Introduction to Information Retrieval for Software Reuse. In *Advances in Software Engineering and Knowledge Engineering,* Vol2, edited by: B. Abriola and G. **Tortor.** World Scientific Publications, Singapore,1993.

[8] J. Poulin and K. Yglesias. Experiences with a faceted July 1987, pp. 129-137. Classification Scheme in a Large Reusable Software Library (RSL). In *The Seventh Annual International Computer Software and Applications Conference (COMPSAC'93),* 1993, pp. 90-99.

[9] W. Frakes and T. Pole. *An Empirical Study **of** Representation Methods for Reusable Software Components.* IEEE Transactions on Software Engineering, August 1994, pp. 617- 630.

[10] R. Prieto-Diaz and P. Freeman. *Classifving Software for Reusability.* IEEE Software, January 1987, pp. 6-16.

[11] R. Prieto-Diaz. *Implementing Faceted Classification for Software Reuse.* Communications of The ACM May

[12] *The Chambers Diction,ary.* Chambers Harrap Publishers Ltd, 1993.

[13] J. F. Koegel Buford. *Uses of Multimedia Information* Multimedia Systems, ACM Press (edited by J. F. Koegel Buford), 1994.

[14] R.Heller and C.D. Martin. *A Media Taxonomy,*IEEE Multimedia, Winter 1995, V01.2, No.4, **pp.** 36 - 45.

[15] S.Feldman, *Make - A P,rogram for Maintaining Computer Programs,* Software - Practice and Experience.

[16] W. F. Tichy. *RCS - A System **for** Version Control.* Software - Practice and Experience, July 1985, pp. 637-654.

[17] B.O'Donovan, J.B.Grimson. *A Distributed Version Control-System For Wide Area Networks* Software Engineering Journal, 1990, vo1.5, no.5, pp. 255-262

[18] A. Dix, T. Rodden and I. Sommerville. *ModellingVersions in Collaborative work.* IEE Proceedings on Software Engineering, Vol. 144, No.4, August 1997, pp.

[19] M.Hanneghan, M.Merabti and G. Colquhoun. *A Viewpoint Analysis Reference Model for Concurrent Engineering.* Accepted for Computers in Industry, June 1998.

[20] J. Plaice and W. Wadge. *A New Approach to Version Control.* IEEE Transactions on Software Engineering.

[21] P. Chen, R. Hennicker and M. Jarke. *On the Retrieval of Reusable Software Components.* IEEE Transactions on Software Engineering, September 1993.

[22] M. Fugini and **S.** Faustle. *Retrieval of Reusable Components in a Development Ififormation System.* IEEE Transactions on Software Engineering, September 1993.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Application of Honeypots to study character of attackers based on their accountability in the network

Tushar Kanti,
Department Of Computer Science,
L.N.C.T, Bhopal,India
Kanti0555@gmail.com

Vineet Richhariya,
Head Of Department Of Computer Science,
L.N.C.T, Bhopal,India
vineet_rich@yahoo.com

Vivek Richhariya,
Department Of Computer Science ,
L.N.C.T, Bhopal,India
vivek_rich@yahoo.com

Abstract— **Malware in the form of computer viruses, worms, trojan horses, rootkits, and spyware acts as a major threat to the security of networks and creates significant security risks to the organizations. In order to protect the networked systems against these kinds of threats and try to find methods to stop at least some part of them, we must learn more about their behavior, and also methods and tactics of the attackers, which attack our networks. This paper makes an analysis of observed attacks and exploited vulnerabilities using honeypots in an organization network. Based on this, we study the attackers behavior and in particular the skill level of the attackers once they gain access to the honeypot systems. The work describes the honeypot architecture as well as design details so that we can observe the attackers behavior. We have also proposed a hybrid honeypot framework solution which will be used in the future work.**

Keywords- Honeypot; Accountability; Classification; Honeynet; Virtual Machines; Honeyd

I. INTRODUCTION

A number of tools have been developed to defend against the attacks that organizations are facing during the recent past. Firewalls, for example, help to protect these organizations and prevent attackers from performing their activities. Intrusion Detection Systems (IDS) are another example of such tools allowing companies to detect and identify attacks, and provide reaction mechanisms against them, or at least reduce their effects. But these tools sometimes lack functionality of detecting new threats and collection of more information about the attacker s activities, methods and skills. For example, signature based IDS s are not capable of detecting new unknown attacks, because they do not have the signatures of the new attacks in their signature database. Thus, they are only able to detect already known attacks. Nevertheless, in order to better protect an organization and build efficient security systems, the developers should gain knowledge of vulnerabilities, attacks and activities of attackers. Today many non-profit research organizations and educational institutions research and analyze methods and tactics of the so-called blackhat community, which acts against their networks. These organizations usually use honeypots to analyze attacks and vulnerabilities, and learn more about the techniques, tactics, intention, and motivations of the attackers [7]. The concept of honeypots was first proposed in Clifford Stoll's book "The Cuckoo's Egg", and Bill Cheswick's paper "An Evening with Berferd"[8]. A Honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource. Honeypots are classified into three types [6]. The first classification is according to the use of honeypots, in other word for what purpose they are used: production or research purpose. The second classification is based on the level of interactivity that they provide the attackers: low or high interaction honeypots. The last one is the classification of honeypots according to their implementation: physical and virtual honeypots. Honeypots as an easy target for the attackers can simulate many vulnerable hosts in the network and provide us with valuable information of blackhat community. Honeypots are not the solution to the network security, they are tools which are implemented for discovering unwanted activities on a network. They are not intrusion detectors, but they teach us how to improve our network security or more importantly, teach us what to look for. Another important advantage of using honeypots is that they allow us to analyze how the attackers act for exploiting of the system's vulnerabilities. The goal of our paper is to study the skill level of the attackers based on their accountability in the honeypot environment. In this paper, we provide the vulnerable systems for the attackers which are built and set up in order to be hacked. These systems are monitored closely, and the attackers skills are studied based on the gathered data.

In order to react properly against detected attacks, the observed skill and knowledge of the attackers should be taken into account when the counter measure process is activated by the security system designers. Therefore, the experimental studies of the attacker's skill level would be very useful to design proper and efficient reaction model against the malwares and blackhat community in the organization's computer network.

The work presented in this paper creates the following main contributions to help learning the attacker s skill level:

Proposing the virtual honeypot architecture and proposing an improved hybrid honeypot framework.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

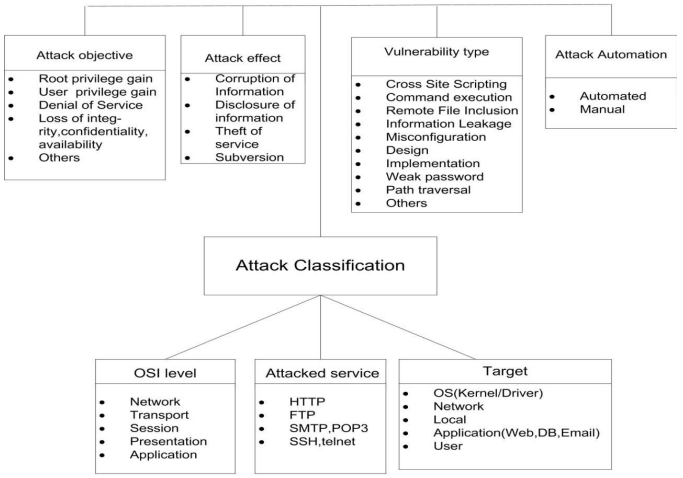
II. BACKGROUND

Based on honeypot techniques researchers have developed many methods and tools for the collection of malicious software. The book [3] and the honeynet project [7], as main sources of our work, provide useful guidelines for the implementation of honeypots and practically experimental tools which have been used in different honeypot projects. Among them there are some honeypot projects which are related to our work. One of the main references which we used often was research outcomes of Leurrecom honeypot project [18]. The Leurrecom project has been created by the Eurocom Institute in 2003. The main goal of this project was to deploy low-interaction honeypots across the internet to collect data and learn more about the attacks which were gathered by their platforms in over 20 countries all over the world. Also we benefited from the research papers of LAAS (The Laboratory of Analysis and Architecture of Systems) [19, 20] for deployment of high-interaction honeypots and precise analysis of the observed attacks, attackers skills and exploited vulnerabilities.

The first time the hybrid honeypot framework has been published in the research paper by Hasan Artail. He proposed this framework [24] in order to improve intrusion detection systems and extend the scalability and flexibility of the honeypots. This approach was helpful when we designed our own Hybrid Honeypot architecture which will be proposed as a future work.

There are two important taxonomies on attack processes: Howard s computer and network security taxonomy [33] and Alvarez s Web attacks taxonomy [43]. Howard s taxonomy classifies the whole attack process of an attacker. The other taxonomy also focus on the attack process, thus it is based on the attack life cycle in analysis of Web attacks. There is also a taxonomy proposed by Hansman and Hunt s [36] which has a four unique dimensional taxonomy that provide a classification covering network and computer attacks. The paper of Wael Kanoun et al. [44] describes the assessment of skill and knowledge level of the attackers from a defensive point of view. Tomas Olsson s work [45] discusses the required exploitation skill-level of the vulnerability and the exploitation skill of the attacker which are used to calculate a probability estimation of a successful attack. The statistical model created by him is useful in order to incorporate real-time monitor data from a honeypot in assessing security risks. He also classifies exploitation skill-levels into Low, MediumLow, MediumHigh, and High levels.

Once attacks, vulnerabilities have been identified, analyzed and classified, we also need to study the exploitation skill of the attackers. We notice that each attacker is a part of the attacker community, and thus, we do not study them individually in the terms of skill level, but as a group. Every attacker has a certain amount of skills and knowledge according to difficulty degree of the exploitation of the vulnerabilities which he has gained access to. The complexity score is based on the difficulty of the vulnerability exploitation, and thus, it also allows us to learn how the attackers are skilled when they successfully exploit the vulnerabilities of our honeypots [39].



"Fig.1" Attack classification

Table 1 Comparison between Honeypots

High-interaction honeypot	Low-interaction honeypot		Hybrid honeypot
- Slow	+ Fast		+Fast
+ Able to detect unknown attacks + 0 False positive	- Unable to detect unknown attacks	→	+ Able to detect unknown attacks + 0 False positive
- Unable to deal with time bombs and user interaction	+ Able to deal with time bombs and user interaction		+ Able to deal with time bombs and user interaction
- Expensive	+ Cheap		+ Better RoI
- Difficult to setup and operate	+ Easy to setup and operate		- Difficult to setup and operate

III. METHOD

We decided to deploy both low and high-interaction honeypots in our experiment. This permitted us to provide comprehensive statistics about the threats, collect high-level information about the attacks, and monitor the activities carried out by different kind attackers (human beings, automated tools).This paper presents the whole architecture used in our work and propose a hybrid honeypot framework that will be implemented in the future.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In the hybrid honeypot system, low-interaction honeypots play the role of a gateway to high-interaction honeypots. Low-interaction honeypots filter out incoming traffic and provide the forwarding of selected connections. In other words, a low-interaction honeypot works as proxy between attacker and the high-interaction honeypot. Hybrid systems include scalability of low interaction honeypots and fidelity of high interaction honeypots [24]. In order to achieve this, low interaction honeypots must be able to collect all of the attacks while unknown attacks should be redirected to high-interaction honeypots. Attackers without any restrictions can get access to high-interaction honeypots which have high fidelity. By using a hybrid architecture, we can reduce the cost of deploying honeypots. But due to lack of time we did not implement the proposed hybrid honeypot architecture.

IV. PROPOSED ARCHITECTURE DETAILS

For our experiment, we designed a honeypot architecture which combines the both low and high interaction honeypots as shown in [Fig 1]. For the low-interaction part we can use Honeyd [2] and for the high-interaction part we can use a virtual honeynet architecture based on the Virtualbox virtualization software [13]. Honeyd is a framework for virtual honeypots that simulates virtual computer systems at the network level. It is created and maintained by Niels Provos [10]. This framework allows us to set up and run multiple virtual machines or corresponding network services at the same time on a single physical machine. Thus, Honeyd is a low-interaction honeypot that simulates TCP, UDP and ICMP services, and binds a certain script to a specific port in order to emulate a specific service. According to the following Honeyd configuration template we have a windows virtual honeypot which is running on 193.x.x.x IP address. This "Windows" template presents itself as Windows 2003 Server Standard Edition when an attacker wants to fingerprint the honeypot with NMap or XProbe.

create windows

set windows personality "Windows 2003 Server Standard Edition"

add windows tcp port 110 "sh scripts/pop3.sh"

bind windows 193.10.x.x

When a remote host connects to TCP port 110 of the virtual Windows machine, Honeyd starts to execute the service script *./scripts/pop3.sh.* There are three honeynet architectures which have been developed by the Honeynet alliance [7]

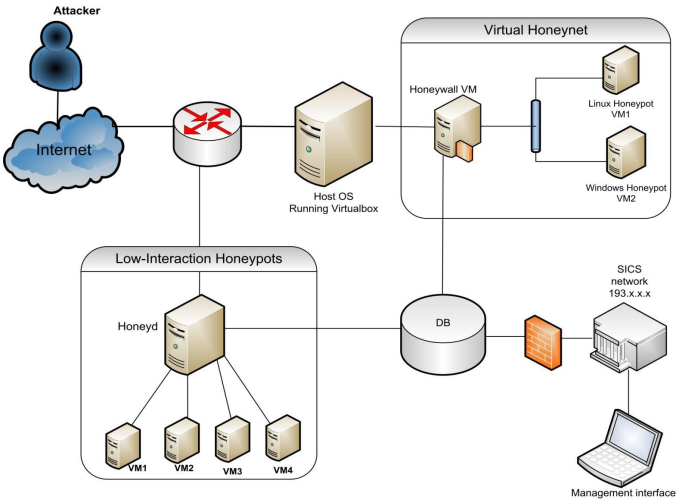
GEN I

GEN II

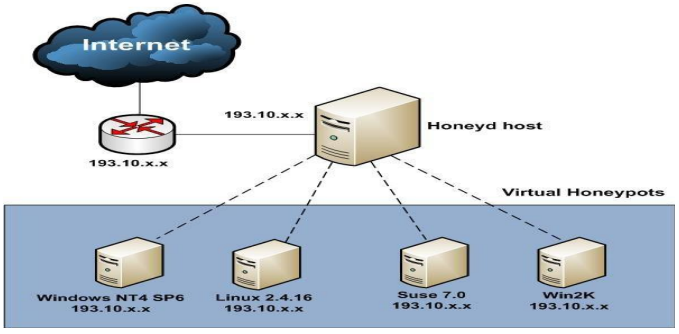
GEN III

GEN I was the first developed architecture and had limited functionality in Data Capture and Data Control. In 2002, GEN II Honeynets were developed in order to address the issues with GEN I Honeynets, and after two years, GEN III was released. GEN II 15 and GENIII honeynets have the same architecture. The only difference between them is the addition of a Sebek server [25] installed in the honeywall within GEN III architecture. The low- and high-interaction honeypots are deployed separately, and the backup of the collected attack data on each host machine of the low and high-interaction honeypots is stored in a common database on a remote machine.

In our design, we used only two physical machines which contain the virtual honeypots and a remote management machine to remotely control the collection of attack data and to monitor the activities and processes on the honeypots. All of the honeypots are deployed and configured on the virtual machines. Using virtualization can help them replace their servers with virtual machines on a single physical machine. Some organizations have been developing their own virtualization solutions which many of them are free and open source.

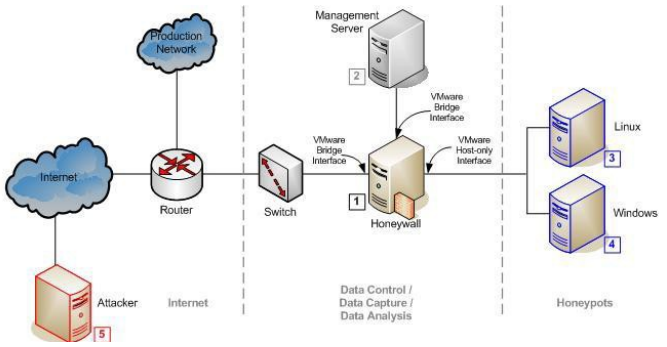


"Fig.2" Proposed Architecture



"Fig.3" Honeyd Framework

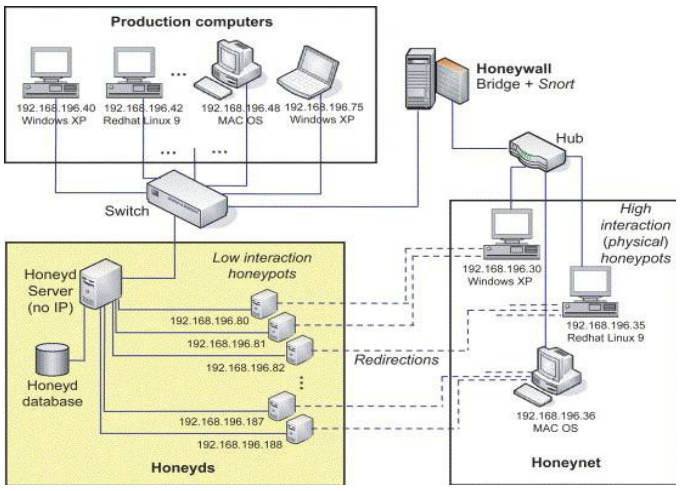
http://sites.google.com/site/ijcsis/
ISSN 1947-5500



"Fig.4" GEN III Honeynet architecture

V. PROPOSED HYBRID HONEYPOT FRAMEWORK (FUTURE WORK)

As a future work we propose an improved hybrid honeypot framework. We already mentioned above that, the first time hybrid honeypot framework has been proposed by Hasan Artail [24]. The hybrid honeypot framework is shown in "Fig.5". It consists of one single common gateway for external traffic and three different internet zones. Production server and clients are in the first zone. The second zone consists of Honeyd server. The Honeyd server has three different services. The first one is for collecting incoming traffic, and stores them in the Honeyd database. The second service generates honeypots based on the statistics provided by the database [24] and the third service provides redirection between low and high interaction honeypots. The last zone consists of an array of high-interaction honeypots running on Physical Machines. As we can see, by default, all the connections are directed into the second zone. And the redirection can happen where the low interaction honeypot filters the traffic to a high interaction honeypot in the third zone. This kind of method can prevent attackers from identifying the existence of the honeypot environment, and provides better configuration to monitor attacks in detail.



"Fig.5" Hybrid Honeypot Framework

VI. CONCLUSION

In this paper, a honeypot architecture is proposed and being used for gathering attack data and tracking the activities carried out by the attackers. We can analyze and classify the observed attacks and vulnerabilities. The aim is to study the attackers skill and knowledge based on this analysis We are successful in this task. It appears that most of the observed attacks are automated and carried out by script kiddies. We can identify different types of attackers based on the nature of their attack.
I hope that this work will help organizations to select proper protection mechanism for their networks by evaluating the impact of detected attacks, and taking into consideration the attacker's skill and knowledge level.

As a future work, We have proposed an improved hybrid honeypot architecture with a different approach to collecting attack data and learning about the attackers skills. By using a hybrid architecture, we can reduce the cost of deploying honeypots. Thus, it will prove to be fruitful for different organizations.

REFERENCES

[1] M. Jakobsson and Z. Ramzan. Crimeware: Understanding New Attacks and Defenses. Addison-Wesley Professional, 2008.

[2] Honeyd, http://www.honeyd.org/

[3] Virtual Honeypots: From Botnet Tracking to Intrusion Detection 2007
by Niels Provos; Thorsten Holz

[4] Conceptual framework for a Honeypot solution
Christian Döring, M.Sc.University of Applied Sciences Darmstadt, Department of Informatics (FHD)

[5] A Guide to Different Kinds of Honeypots
http://www.securityfocus.com/infocus/1897

[6] Lance Spitzner.Honeypots, Definitions and Value of Honeypots .
http://www.spitzner.net May, 2002

[7] The Honeynet Project."Know your enemy" (http://project.honeynet.org).

[8] Clifford Stoll.The Cuckoo s egg. ISBN: 0743411463

[9] http://en.wikipedia.org/wiki/Honeypot_(computing)

[10] Niels Provos. A virtual honeypot framework. In Proceedings of 13th USENIX Security Symposium, pp. 1–14. USENIX, 2004.

[11] Lance Spitzner . Honeypots: Tracking hackers Addison Wesley Professional, September 2002

[12] Nepenthes. http: //nepenthes.mwcollect.org

[13] SUN Microsystems. VirtualBox. http://www.virtualbox.org/.

[14] "Know Your Enemy: Honeywall CDROM Roo",
http://old.honeynet.org/papers/cdrom/roo/index.html

[15] Honeypotting with VMware - basics
http://seifried.org/security/ids/20020107-honeypot-vmware-basics.html
[16] The Value of Honeypots, Part One: Definitions and Values of Honeypots

by Lance Spitzner with extensive help from Marty Roesch last updated October 10, 2001
http://www.securityfocus.com/infocus/1492

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[17] F.Pouget,M.Dacier,LEURRE'COM:The Eurocom Honeypot Project 64

[18] [Kaâniche et al. 2006] M.Kaâniche, E.Alata, V.Nicomette, Y.Deswarte, M.Dacier, Empirical Analysis and Statistical Modeling of Attack Processes based on Honeypots, 25-28 June 2006

[19] Alata, Eric;Nicomette, V;Kaâniche, M;Dacier, Marc;Herrb, M
Lessons learned from the deployment of a high-interaction honeypot

[20] A Hybrid Honeypot Architecture for Scalable Network Monitoring
Michael Bailey, Evan Cooke, David Watson, Farnam Jahanian Niels Provos
University of Michigan October 27, 2004

[21] Hybrid Honeypot System for Network Security
Kyi Lin Lin Kyaw, Department of Engineering Physics, Mandalay Technological University
[22] Advanced Honeypot Architecture for Network Threats Quantification ,Robin G. Berthier 2009

[23] Know your enemy: Web Application Threats
http://www.honeynet.org/papers/webapp/

[24] A hybrid honeypot framework for improving intrusion detection systems in protecting organizational networks. Hassan Artail

[25] Honeynet Project, Sebek, Honeynet Project website
http://project.honeynet.org/tools/sebek/

[26] Shuja, F. (October, 2006). Virtual Honeynet: Deploying Honeywall using VMware. Available: http://www.honeynet.pk/honeywall/index.htm. Last accessed June, 2008.

[27] Know Your Enemy: Defining Virtual Honeynets
http://old.honeynet.org/papers/virtual/

[28] Psacct utility
http://linux.maruhn.com/sec/psacct.html

[29] SMF
http://www.simplemachines.org/

[30] Joel Weise and Brad Powell. Using computer forensics when investigating system attacks.
Sun BluePrints OnLine, Sun Client Solutions Security Expertise Center, April 2005.

[31] Phillipine Honeypot project
http://www.philippinehoneynet.org/

[32] ProjectHoneypot http://www.projecthoneypot.org

[33] J. Howard and T. Longstaff. A common language for computer security incidents. Sandia
Intelligence Labs, 1998.

[34] Lough, Daniel. "A Taxonomy of Computer Attacks with Applications to Wireless Networks," PhD thesis, Virginia Polytechnic Institute and State University, 2001.

[35] Lindqvist U, Jonsson E. How to systematically classify computer security intrusions. IEEE Security and Privacy 1997:154e63.

[36] Hansman, S., Hunt R., "A taxonomy of network and computer attacks". Computer and Security (2005).

[37] Common Vulnerabilities and Exposures (CVE) http://cve.mitre.org/

[38] National Vulnerability Database http://nvd.nist.gov/

[39] Forum of Incident Response and Security Teams (FIRST). Common Vulnerabilities Scoring System (CVSS). http://www.first.org/cvss/.

[40] MITRE Common Weakness Enumeration http://cwe.mitre.org/

[41] M.A. McQueen et al., "Time-to-Compromise Model for Cyber Risk Reduction Estimation", Quality of Protection: Security Measurements and Metrics, Springer, 2005.

[42] Paulauskas N, Garsva E. Attacker skill level distribution estimation in the system mean time-to-compromise

[43] G. Álvarez, S. Petrović, 'A new taxonomy of web attacks suitable for efficient encoding,' Computers and Security, 22(5), pp. 435-449, 2003.

[44] Automated Reaction based on Risk Analysis and Attackers Skills in Intrusion Detection Systems (2009) Wael Kanoun, Nora Cuppens-boulahia, Frédéric Cuppens

[45] Olsson, Tomas (2009) Assessing Security Risk to a Network Using a Statistical Model of Attacker Community Competence. In: Eleventh International Conference on Information and Communications Security (ICICS 2009), 14-17 Dec 2009, Beijing, China.

AUTHOR'S PROFILE

"Mr. Tushar Kanti is Mtech in Computer Science and Engg. from Laxmi Narayan College Of Technology,Bhopal,INDIA"

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Performance of Input and Output Selection Techniques on Routing Efficiency in Network-On-Chip

Mohammad Behrouzian Nejad
Young Researchers Club, Dezfoul Branch
Islamic Azad University
Dezfoul, Iran
Mohamad.behrouzian@gmail.com

Amin Mehranzadeh
Department of Computer Engineering
Dezfoul Branch, Islamic Azad University
Dezfoul, Iran
Mehranzadeh@iaud.ac.ir

Mehdi Hoodgar
Department of Computer Engineering
Dezfoul Branch, Islamic Azad University
Dezfoul, Iran
Hoodgar@iaud.ac.ir

***Abstract*— Network-On-Chip (NOC) is a new paradigm to make the interconnections inside a System-On-Chip (SOC) system. Networks-On-Chip have emerged as alternative to buses to provide a packet-switched communication medium for modular development of large Systems-On-Chip. The performance of Network-On-Chip largely depends on the underlying routing techniques. Routing algorithm can be classified into three categories, namely, deterministic routing, oblivious routing and adaptive routing. Each routing algorithm has two constituencies: output selection and input selection. In this paper we discuss about some input and output selection techniques which used by routing algorithms. Also, to provide new and more efficient algorithms we examine the strengths and weaknesses of the algorithm.**

Keywords: Network, System-On-Chip, Network-On-Chip, Routing algorithm, Input selection, Output selection.

I. INTRODUCTION

As technology scales and chip integrity grows, on chip communication is playing an increasing dominant role in System-On-Chip design. System-On-Chip complexity scaling driven by the effect of Moore's Law in Integrated Circuits are required to integrate from dozens of cores today to hundreds of cores within a single chip in the near future. The NOC approach has been recently proposed for efficient communication in SOC designs. In order Network-On-Chip is a new paradigm for System on Chip design. Increasing integration produces a situation where bus structure, which is commonly used in SOC, becomes blocked and increased capacitance poses physical problems. Traditional bus in NOC architecture is replaced with a network which is a lot similar to the Internet. Data communications between segments of chip transferred through the network. In the most commonly found organization, a NOC is a set of interconnected switches, with IP cores connected to these switches. NOCs present better performance, bandwidth, and scalability than shared busses [1-8].

II. NETWORK-ON-CHIP

The idea of NOC is derived from large scale computer networks and distributed computing. The Network-On-Chip architecture provides the communication infrastructure for the resources. In this way it is possible to develop the hardware of resources independently as standalone blocks and create the NOC by connecting the blocks as elements in the network. Moreover, the scalable and configurable network is a flexible platform that can be adapted to the needs of different workloads, while maintaining the generality of application development methods and practices. Fig.1 shows a mesh-based NOC, which consists of a grid of 16 cores. Each core is connected to a switch by a network interface. Cores communicate with each other by sending packets via a path consisting of a series of switches and inter-switch links. The NOC contains the following fundamental components [9-13].

a) Network adapters implement the interface by which cores (IP blocks) connect to the NOC. Their function is to decouple computation (the cores) from communication (the network).

b) Routing nodes route the data according to chosen protocols. They implement the routing strategy.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

c) Links connect the nodes, providing the raw bandwidth. They may consist of one or more logical or physical channels.

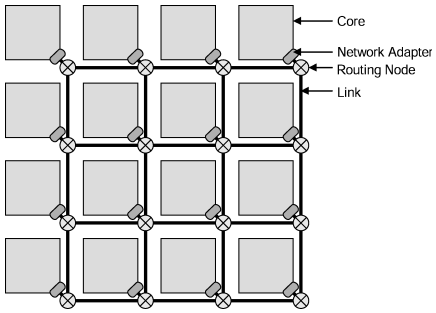


Figure 1. The typical structure of a 4*4 NOC

B. Topology in Network-on-Chip

The job of the network is to deliver messages from their source to their designated destination. This is done by providing the hardware support for basic communication primitives. A well-built network, as noted by Dally and Towles [14], should appear as a logical wire to its clients. An on-chip network is defined mainly by its topology and the protocol implemented by it. Topology concerns the layout and connectivity of the nodes and links on the chip. Protocol dictates how these nodes and links are used [12, 13]. In order Topology determines how the nodes in the network are connected with each other. In a multiple-hop topology, packets may travel one or more intermediate nodes before arriving at the target node. Regular multiple-hop topologies such as mesh and torus are widely used in NOCs. We can use different topologies for the optical data transmission network and the electronic control network respectively [15, 16]. Fig.2 shows some kinds of topology which used in NOC.

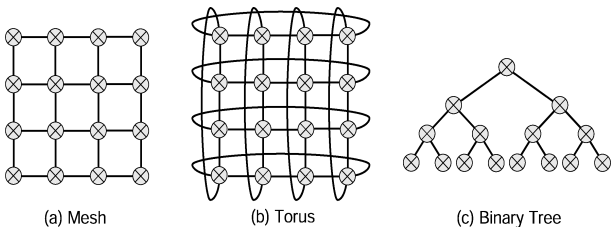


Figure 2. (a) 4-ary 2-cube mesh, (b) 4-ary 2-cube torus and (c) binary tree

C. Routing Algorithms

Routing on NOC is similar to routing on any network. The routing techniques for NOC have some unique design considerations besides low latency and high throughput. Due to tight constraints on memory and computing resources, the routing techniques for NOC should be reasonably simple [5, 6, and 9]. The routing algorithm determines the routing paths the packets may follow through the network graph. It usually restricts the set of possible paths to a smaller set of valid paths. In terms of path diversity and adaptively, routing algorithm can be classified into three categories, namely, deterministic routing, oblivious routing and adaptive routing. Deterministic routing chooses always the same path given the source node and the destination node. It ignores the network path diversity and is not sensitive to the network state. This may cause load imbalances in the network but it is simple and inexpensive to implement. Besides, it is often a simple way to provide the ordering of packets. Oblivious routing, which includes deterministic algorithms as a subset, considers all possible multiple paths from the source node to the destination node, for example, a random algorithm that uniformly distributes traffic across all of the paths. But oblivious algorithms do not take the network state into account when making the routing decisions. The third category is adaptive routing, which distributes traffic dynamically in response to the network state. The network state may include the status of a node or link, the length of queues, and historical network load information [17, 18]. In the NOC, to route packets through the network, the switch needs to implement a routing technique [9]. A routing technique witch used in routing algorithms has two constituencies: output selection and input selection which describes in section D and E.

D. Input Selection Technique

Multiple input channels may request simultaneously the access of the same output channel, e.g., in fig.3 packets p0 of input_0 and p1 of input_1 can request output_0 at the same time. The input selection chooses one of the multiple input channels to get the access. Two input selections have been used in NOC, first-come-first-served (FCFS) input selection and round-robin input selection. In FCFS, the priority of accessing the output channel is granted to the input channel which requested the earliest. Round-robin assigns priority to each input channel in equal portions on a rotating basis. FCFS and round-robin are fair to all channels but do not consider the actual traffic condition [9].

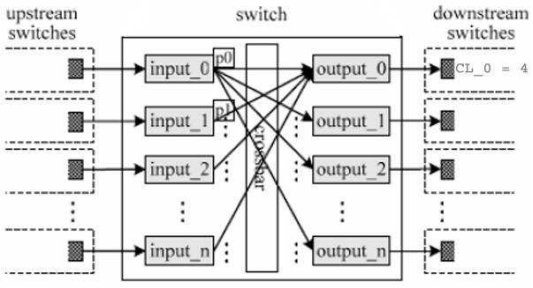


Figure 3. Block diagram of switch in NOC

Dong Wu in [9] a new input selection technique presented which based on Contention Aware Input Selection (CAIS). The main idea behind CAIS is that when two or more input packets both desire the same output channel, the decision as to which packet should obtain the output is made based on upstream contention information. The aim of CAIS is to use contention information to alleviate congestion [9, 19].

In order, the basic idea of CAIS is to give the input channels different priorities of accessing the output channels. The priorities are decided dynamically at run-time, based on

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the actual traffic condition of the upstream switches. More precisely, each output channel within a switch observes the contention level (the number of requests from the input channels) and sends this contention level to the input channel of the downstream switch, where the contention level is then used in the input selection. When multiple input channels request the same output channel, the access is granted to the input channel which has the highest contention level acquired from the upstream switch. This input selection removes possible network congestion by keeping the traffic flowing even in the paths with heavy traffic load, which in turn improves routing performance. Fig. 4 shows the algorithm of CAIS [9]. In CAIS an input channel which has lower CL continuously competing with channels which have higher CL, obviously will be defeated any time. The packets in this channel won't be able to get their required output channel and face with starvation and this will cause the problem of decreasing network efficiency. Thus, there is a starvation possibility in this new input selection technique, because it performs input selection only based on the highest contention level (CL) and the channels with low CL have a little chance for winning. So this input selection technique improved in [20], which in addition to CL, another parameter with the name of AGE for every input channel is taken into consideration and measure of priority will be a compound of CL+AGE. In this technique, the problem of starvation has been resolved.

Contention-Aware Input Selection (CAIS)

req_0..n	request signals from the input channels
out_cl_i	CL of the i^{th} output channel
in_cl_j	CL of the j^{th} input channel acquired from the upstream switch
max_cl	maximum contention level
sel_i	selection signal of the i^{th} output channel
i = 0, 1, 2, ..., n;	*j = 0, 1, 2, ..., n;*

```
01  process observe_cl(req_0..n)
02  begin
03      out_cl_i <= number of request to the ith output channel;
04  end process observe_cl
05
06  process select_input
07  begin
08      max_cl := 0;
09      for all requests loop
10          if in_cl_j >= max_cl then
11              max_cl := in_cl_j;
12              sel_i <= j;
13          end if
14      end loop
15  end process select_input
```

Figure 4. Pseudo VHDL code of the CAIS algorithm

E. Output Selection Technique

A packet coming from an input channel may have a choice of multiple output channels, e.g., in fig.2 a packet p0 of input_0 can be forwarded via output_0, output_1 and so on. The output selection chooses one of the multiple output channels to deliver the packet. Several switch architectures have been developed for NOC [5, 9, and 10], employing XY output selection and wormhole routing. The routing technique proposed in [21] acquire information from the neighboring switches to avoid network congestion and uses the buffer levels of the downstream switches to perform the output selection. A routing scheme which is based on Odd-Even routing algorithm[10] and combines deterministic and adaptive routing is proposed in [22], where the switch works in deterministic mode when the network is not congested, and switches to adaptive mode when the network becomes congested. In the IV, V and VI we describes some kinds of output selection techniques of deterministic routing, oblivious routing and adaptive routing which presented for NOC.

III. IMPORTANT PROBLEMS IN ROUTING ALGORITHMS

Many properties of the NOC are a direct consequence of the routing algorithm used. Among these properties we can cite the following [23]:

a) Connectivity: Ability to route packets from any source node to any destination node.

b) Adaptively: Ability to route packets through alternative paths in the presence of contention or faulty components.

c) Deadlock and live lock freedom: Ability to guarantee that packets will not block or wander across the network forever.

d) Fault tolerance: Ability to route packets in the presence of faulty components. Although it seems that fault tolerance implies adaptively, this is not necessarily true. Fault tolerance can be achieved without adaptively by routing a packet in two or more phases, storing it in some intermediate nodes.

A good routing algorithm should be avoidance from deadlock, live lock, and starvation. Deadlock may be defined as a cyclic dependency among nodes requiring access to a set of resources, so that no forward progress can be made, no matter what sequence of events happens. Live lock refers to packets circulating the network without ever making any progress towards their destination. Starvation happens when a packet in a buffer requests an output channel, being blocked because the output channel is always allocated to another packet [7, 20, and 23].

IV. DETERMINISTIC ROUTING ALGORITHMS

Many properties of the NOC are a direct consequence of the routing algorithm used. The XY algorithm is deterministic. Flits are first routed in the X direction, until reaching the Y coordinate, and afterwards in the Y direction. If some network hop is in use by another packet, the flit remains blocked in the switch until the path is released [5, 7].

V. OBLIVIOUS ROUTING ALGORITHMS

A. Dimension Order Routing

This routing algorithm routes packets by crossing dimensions in increasing order, nullifying the offset in one dimension before routing in the next one. A routing example is shown in Fig.5 Note that dimension-order routing can be executed at the source node, storing information about turns

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(changes of dimension) in the header [6]. This is the street-sign routing algorithm described above. Dimension-order routing can also be executed in a distributed manner. At each intermediate node, the routing algorithm supplies an output channel crossing the lowest dimension for which the offset is not null.

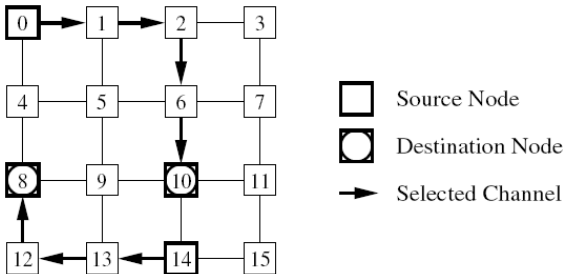


Figure 5. Routing example for dimension-order routing on a 2-D mesh

B. O1TURN Routing Algorithm

An oblivious routing algorithm (O1TURN) for 2-D mesh networks has been described in [24]. O1TURN performs well in the three main criteria as defined in their paper – minimizing number of hops, delivering near optimal worst-case and good average-case throughput, and allowing a simple implementation to reduce router latency. According to the authors, existing routing algorithms optimize some of the above mentioned design goals while sacrificing the others. The proposed O1TURN (Orthogonal One-TURN) algorithm addresses all three of these issues. O1TURN allows each packet to traverse one of two dimension-ordered routes (X first or Y first) by randomly selecting between the two options. It is an interesting 2-D extension to the Randomized Local Balanced routing (RLB) algorithm utilized in ring topologies [6].

C. ROMM Routing Algorithm

ROMM is a class of Randomized, Oblivious, Multi-phase, Minimal routing algorithms [25]. For a large range of traffic patterns ROMM is superior to DOR since it allows minimal routing with some load balancing. ROMM randomly chooses an intermediate node in the minimal rectangle between the source and destination nodes, and then routes packets through the intermediate node using DOR. The simplicity and good average-case performance of ROMM make it a desirable algorithm for systems where average-case throughput is important. However, ROMM fails to provide good worst-case throughput since source/destination pairs can create additional congestion in channels not in the row and column of source and destination nodes. Although the worst-case throughput is undesirably low, in practice it does not occur very frequently. In fact people were generally unaware of the exact worst case traffic pattern until an analytical approach 4 for calculating worst case throughput was described in [6]. Therefore, ROMM is a popular choice for networks where the worst-case throughput is not critical.

D. VALIANT Routing Algorithm

The VALIANT routing algorithm guarantees optimal worst-case throughput by randomizing every traffic pattern [26]. VALIANT randomly picks an intermediate node from any node in the network and routes minimally from source to intermediate node and then from the intermediate to the destination node. This is a non-minimal routing algorithm which destroys locality and hurts header latency, but guarantees good load balancing. It can be used if the worst-case throughput is the only critical measure for the network. IVAL (Improved Valiant's randomized routing) is an improved version of the oblivious Valiant's algorithm. It is a bit similar to turn around routing. On the algorithms first stage packets are routed to a randomly chosen point between the sender and the receiver by using an oblivious dimension order routing. The second stage of the algorithm works almost equally, but this time the dimensions of the network are gone through in reversed order. Deadlocks are avoided in IVAL routing by dividing router's channels to virtual channels. Full deadlock avoidance requires a total of four virtual channels per one physical channel.

VI. ADAPTIVE ROUTING ALGORITHMS

A. Q-Routing

The functionality of a Q-routing algorithm is based on the network traffic statistics. The algorithm collects information about latencies and congestions and maintains statistics about network traffic. The Q-routing algorithm does the routing decisions based on these statistics [27, 28].

B. Odd-Even Routing Algorithm

The odd-even adaptive routing algorithm was proposed by Chiu [10]. In his paper on the odd-even turn model. The model shows how selectively restricting the directions routing turns are permitted to take provides the resource ordering needed to ensure that the routing algorithm remains deadlock free. The odd-even routing algorithm prohibits even column routing tiles from routing east to north and east to south while prohibiting odd column routing tiles from routing north to west and south to west. Among adaptive routing algorithms without virtual channel support [7], the odd-even scheme routes in a more evenly distributed fashion across the network. A minimal route version of odd-even was selected to ensure the network doesn't live lock and also to minimize energy consumption.

C. DyAD Routing Algorithm

The acronym DyAD stands for: Dynamically switching between Adaptive and Deterministic routing modes. The intention of the DyAD routing scheme Hu [22] is to propose a new paradigm for the design of a Network-On-Chip router that allows the NOC routing algorithm to exploit the advantages of both deterministic and adaptive routing. As such, DyAD is presented as a hybrid routing scheme that can perform either

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

adaptive or deterministic routing to achieve best possible throughput. With the DyAD hybrid routing scheme, the network continuously monitors its local network load and makes the choice of whether to use an adaptive or deterministic routing mode based on local network load. When the network is not congested a DyAD router works in a deterministic mode and thus can route with the low latency that is facilitated by deterministic routing. When the network becomes congested, a DyAD router switches to routing in adaptive mode to avoid routing to congested links by exploiting other less congested routes. The authors implemented one possible variation of the DyAD hybrid scheme that employs two flavors of the odd-even routing scheme, one flavor as a deterministic scheme and one flavor as an adaptive routing scheme. By measuring how full local FIFO queues are, a router may switch between deterministic and adaptive modes. Further, the DyAD scheme proposed is shown to be deadlock and live lock free in the presence of the mixture of deterministic and adaptive routing modes. Performance measurements are reported that highlight the advantages of this hybrid approach. Measurements are reported for several permutation traffic patterns as well as a real world multimedia traffic pattern. Evidence is presented that the additional resources required to support a hybrid routing scheme are minimal.

D. Hot-Potato Routing

The hot-potato routing algorithm routes packets without temporarily storing them in router's buffer memory. Packets are moving all the time without stopping before they reach their destination. When one packet arrives to a router, the router forwards it right away towards packet's receiver but if there are two packets going to same direction simultaneously, the router directs one of the packets to some other direction. This other packet can flow away from its destination. This occasion is called misrouting. In the worst case, packets can be misrouted far away from their destination and misrouted packets can interfere with other packets. The risk of misrouting can be decreased by waiting a little random time before sending each packet. Manufacturing costs of the hot-potato routing are quite low because the routers do not need any buffer memory to store packets during routing [6, 29].

E. 2TURN

2TURN algorithm itself does not have an algorithmic description. Only algorithms possible routing paths are determined in a closed form. Routing from sender to receiver with 2TURN algorithm always consists of 2 turns that will not be U-turns or changes of direction within dimensions. Just as in the IVAL routing, a 2TURN router can avoid deadlock if all router's physical channels are divided to four virtual channels [6].

VII. CONCLUSIONS

Network-On-Chip is a technology of future on System on Chip implementations. Content as can be concluded that the input and output selection techniques which used in routing algorithm, significant impact on Network on Chip performance is better. This paper shows importance of routing algorithm in rate of delays in the routing and network better performance and yet, some of the most popular and efficient routing algorithms which proposed for Network on Chip, introduced and examined. Most existing algorithms, despite significant improvements in reducing the average latency and network performance have improved. But still the more defects and incomplete to improve performance of Network on Chip, it is felt. The paper also examines the strengths and weaknesses of the algorithms, to provide new and more efficient algorithms can be useful. The some outlines and features of the routing algorithms presented above are listed in Table. I.

TABLE I. OUTLINES AND FEATURES OF ROUTING ALGORITHMS[6]

Algorithm	Outlines	Features
XY	routing first in X and then in Y dimension	simple, loads network deadlock-and live lock free
DOR	routing in one dimension at a time	Simple
Q-Routing	Statistics based routing	uses the best path
Odd-Even	Turn model	Deadlock free
DyAD	Dynamically Deterministic and Adaptive mode	uses the best path
2TURN	slightly determined	Efficient
Hot-potato	routing without buffer memories	cheap, sometimes misrouting
IVAL	Improved turnaround routing	Uses efficiently whole network

REFERENCES

[1] O. Tayan, "*Extending Logical Networking Concepts in Overlay Network-on-Chip Architectures*", International Journal of Computer Science and Information Security, Volume 8 No. 1, pp 64-67, 2010.

[2] L. Wang, Y. Cao, X. Li , X. Zhu, "*Application Specific Buffer Allocation for Wormhole Routing Networks-on-Chip*", Network on Chip Architectures(NOCARC), 41st Annual IEEE/ACM International Symposium on Micro architecture (MICRO-41),Italy, 2008.

[3] T.C. Huang, Umit Y. Ogras, R. Marculescu, "Virtual *Channels Planning for Networks-on-Chip*", Proceedings of the 8th International Symposium on Quality Electronic Design (ISQED'07), IEEE, 2007.

[4] W.J. Dally and B. Towles, "*Route Packets, not Wires: On-chip Interconnection Networks*," DAC, June, 2001.

[5] M. Behrouzian-nejad, "*A survey of performance of deterministic & adaptive routing algorithms in network on chip architecture*", Proceedings of 2nd Regional Conference on Electrical & Computer, Naein Branch, Iran, 2011.

[6] V. Rantala, T. Lehtonen, J. Plosila, "*Network on Chip Routing Algorithms*", Turko center for computer science, Joukahaisenkatu 3-5 B, 20520 Turku, Finland, 2006.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7] A.V.de Mello L.C.Ost, F.G.Moraes, N.L.V.Calazans, "*Evaluation of Routing Algorithms on Mesh Based NoCs*", PUCRS, Av.Ipiranga, 2004.

[8] L. Benini, G. De Micheli, "*Networks on chips: a new SoC paradigm*", IEEE Computer, v. 35(1), pp. 70-78, 2002.

[9] Dong Wu, M.Bashir Al-Hashimi, T.Marcus Schmitz," *Improving Routing Efficiency for Network-on-Chip through Contention-Aware Input Selection*", Proceedings of 11th Asia and South Pacific Design Automation Conf. Japan, 2006.

[10] G.-M. Chiu, "*The odd-even turn model for adaptive routing*", IEEE Transactions on Parallel and Distributed Systems, vol. 11, pp.729-38, 2000.

[11] L. M. Ni and P. K. McKinley, "A *survey of wormhole routing techniques in direct networks*", Computer, vol. 26, pp. 62-76, 1993.

[12] A. Jantsch, H. Tenhunen, "Networks *on Chip*", Kluwer Academic Publishers Dordrecht, 2003.

[13] T.BJERREGAARD, S.MAHADEVAN, "A *Survey of Research and Practices of Network-on-Chip*", ACM Computing Surveys, Vol. 38, March 2006.

[14] W. J. DALLY, B. TOWLES, "*Route packets, not wires: On-chip interconnection networks*", Proceedings of the 38th Design Automation Conference (DAC). IEEE, 684–689. 2001.

[15] F. Gebali, H. Elmiligi, M.W.El-kharashi, "*Network on chips theory and practice*", Taylor & Francis Group, 2009.

[16] T.Ye, L. Benini, and G. De Micheli, "*Packetization and routing analysis of on-chip multiprocessor networks*", Journal of Systems Architecture, vol. 50, pp. 81-104, 2004.

[17] Z. Lu, "*Design and Analysis of On-Chip Communication for Network-on-Chip Platforms*", Department of Electronic, Computer and Software Systems School of Information and Communication Technology Royal Institute of Technology (KTH) Sweden, 2007.

[18] W. J. Dally and B. Towles. "*Principles and Practices of Interconnection Networks*", Morgan Kaufman Publishers, 2004.

[19] E.Chron, G.Kishinevsky, B.Nefcy, N.Patil," *Routing Algorithms for 2-D Mesh Network-On-Chip Architectures*" http://www.cva.stanford.edu, 06/15/2011.

[20] E .Behrouzian-Nejad, A. Khademzadeh , "*BIOS:A New Efficient Routing Algorithm for Network on Chip*", journal of Contemporary Engineering Sciences, Vol. 2,no. 1, 37 – 46, 2009.

[21] E. Rijpkema, K. Goossens, A. Radulescu, J. Dielissen, J. Van Meerbergen, P. Wielage, and E. Waterlander, "*Trade-offs in the design of a router with both guaranteed and best-effort services for networks on chip*", IEEE Proceedings: Computers and Digital Techniques, vol. 150, pp. 294-302, 2003

[22] J. Hu, R. Marculescu: "*DyAD – Smart Routing for Networks-on-Chip*", Proceedings, 41st Design Automation Conference, 2004.

[23] J.Douato, s. Yalamanchili, L. Ni, Morgan, "*Interconnection Network*", Elsevier Science, 2003.

[24] Seo, D., A. Ali, W. Lim, N. Rafique and M. Thottethodi, "*Near Optimal Worst-case Throughput 10 Routing for Two-Dimensional Mesh Network*", ISCA 2006.

[25] Nesson, T. and S. L. Johnsson, "*Romm routing on mesh and torus networks," Proc. of the seventh annual ACM symposium on Parallel algorithms and architectures*", ACM Press, pp 275–287, 1995.

[26] Valiant, L. G. and G. J. Brebner, "*Universal schemes for parallel communication*", Proc. thirteenth annual ACM symposium on Theory of computing, pages 263–277. ACM Press, 1981.

[27] M.Majer, C. Bobda, A. Ahmadinia, J. Teich, "*Packet Routing in Dynamically Changing Networks on Chip*" ,Proceedings, 19th IEEE International Parallel and Distributed Processing Symposium, 4–8 April 2005.

[28] R. Dobkin, R. Ginosar, I. Cidon, "*QNoC asynchronous router with dynamic virtual channel allocation*", International Symposium on Networks-on-Chip, 2007.

[29] U. Feige, P. Raghavan, "*Exact Analysis of Hot-Potato Routing*", 33rd Annual Symposiumon Foundations of Computer Science, pp 553–562, 1992.

AUTHORS PROFILE

Mohammad Behrouzian Nejad Was born in Dezful, a city in southwestern of Iran, in 1990. He is currently Active Member of Young Researchers Club (YRC) and a B.Sc student at Islamic Azad University, Dezful Branch, Dezfoul, Iran. His research interests are Computer Networks, Information Technology and Data Mining.

Amin Mehranzadeh was born in Dezfoul, Iran, in 1979. He received a B.Sc. degree in computer architecture from Azad University of Dezful, Khuzestan, Iran in 2002 and a M.Sc. degree in computer architecture systems from Azad university of Markazi, Arak, Iran in 2010. He is currently teaching in the department of Computer Engineering at the Azad University of Dezful, Iran. His research interests include Network on Chip (NOC) and simulation of Routing Algorithms.

Mehdi Hoodgar was born in Dezfoul, Iran, in 1978. He received a B.Sc. degree in computer architecture from Azad University of Dezful, Khuzestan, Iran in 2002 and a M.Sc. degree in computer architecture systems from Azad university of Khuzestan, Dezful, Iran in 2010. He is currently teaching in the department of Computer Engineering at the Azad University of Dezful, Iran.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Critical Analysis of Design Criteria And Reliability Of Safety Instrumented System (Sis) For Offshore Oil & Gas Production Platforms In India

Rakesh Sethi[1], Manjeet Patterh[2]
[1]Superintending Engineer ONGC Research scholar Punjabi university Patiala, India
[2]Director, University College of Engineering Punjabi University Patiala, India

ABSTRACT
In this paper observed that there is a growing need in offshore oil & gas industry to gain insight into the significant aspects and parameters of safety instrumented systems so as to manage the process in a more reliable and safer manner. The diversity of issues and the use of different subsystems demand a multidisciplinary team with expertise in process, instrumentation, control, safety, maintenance, reliability and management to develop the basis for the design, implementation, and maintenance and successfully design Criteria and Reliability of Safety Instrumented System for Offshore Oil & Gas Production Platform in India.

Keywords: safety Instrumented System, Offshore Oil and Gas Industry.

I. INTRODUCTION

As hydrocarbon demand continues to rise, oil and gas companies are forced to explore and exploit at increased water depths, in harsher environments and to handle fluids at higher pressures and temperatures. Offshore process, well-head flow lines, risers, sub-sea pipelines and plant structures are increasing in complexity, warranting more reliable and effective methods of risk assessment and mitigation techniques with minimum possible cost. As a part of overall risk management policy, E&P (Exploration and Production) companies use a variety of safeguards or protection layers to reduce the risk to the tolerable level.

They are devices, systems or actions that are capable of preventing a scenario from proceeding to an undesired consequence. e.g. inherently safe design features, physical protection such as relief devices, post-release physical protection such as fire suppression systems, plant & community emergency response plan, Basic Process Control System (BPCS) and Safety Instrumented System (SIS). Safety Instrumented Systems are probably one of the most important risk reduction and mitigation measures.

Safety Instrumented System (SIS) is a highly reliable system of interconnected sensors, final elements and logic meant to fulfill the intended safeguarding functions of the concerned process. Purpose of the SIS is to take the process to a safe state when predetermined conditions are violated such as set points for pressure, temperature or any other process parameter. It consists of the instrumentation or controls that are installed for the purpose of identification and mitigation of process hazards.

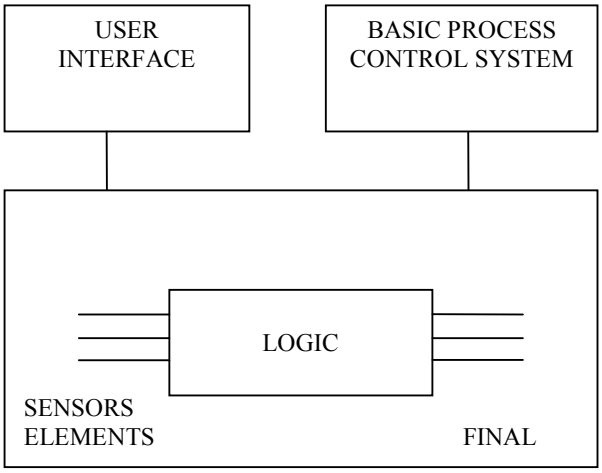


Fig: Definition of safety Instrumented System

To maintain a safe state of process, safety instrumented functions are implemented in SIS and each safety instrumented function is assigned a target safety integrity level (SIL).

SIL is a measure of system reliability in terms of probability of failure of SIS on demand [1]. It is a way to indicate the tolerable failure rate of a particular safety function or in other words, the level of performance needed to achieve the user's process safety objective. Worldwide, within the regulatory framework of country and self defined acceptable risk criteria; companies use various methodologies to determine target SIL for safety instrumented functions of SIS. Methodologies used for determining SIL include, but not limited to modified HAZOP (Hazard & Operability), risk graph, risk matrix, safety layer matrix, layer of protection analysis (LOPA), fault tree analysis (FTA) and Markov Analysis.

Following table shows the relationship between average probability of failure on demand (PFDavg.), availability of the safety system, risk reduction and the SIL levels [2].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Safety Integrity Level (SIL)	Availability	PFDavg.	Risk Reduction
4	0.9999 to 0.99999	10^{-4} to 10^{-5}	10^4 to 10^5
3	0.9990 to 0.99990	10^{-3} to 10^{-4}	10^3 to 10^4
2	0.9900 to 0.99900	10^{-2} to 10^{-3}	10^2 to 10^3
1	0.9000 to 0.99000	10^{-1} to 10^{-2}	10^1 to 10^2

Safety integrity level (SIL) can be considered as a statistical representation of reliability and availability of safety instrumented system (SIS) at the time of process demand and design of SIS plays a major role in it.

II. SIS DESIGN CONSIDERATIONS

Old offshore oil & gas installations in India are designed on the basis of recommended practices mentioned in API RP14C [3], API RP14G [4] and API 14J [5]. When these recommended practices were developed, safety systems were pneumatic or relay based and offshore processes were relatively simple. Time has changed, and so has our need for the right tools. Present requirement is programmable logic controllers with more and more complex logic and standards like IEC 61511 or ANSI ISA S-84 are more relevant for instrumentation of offshore safety . Recommended practices like RP14C were conceived to lower risk associated with personal injury only. They were created to address "dangerous" failures and are not concerned with "safe" failures because they don't lead to personnel injury. Present day safety systems are more integrated with overall risk management of the companies. They are created to minimize dangerous failures, but they also recognize that some safe failures (nuisance trips) are responsible for unnecessary downtime and revenue loss. This increases safety as well as profitability but also calls for "measurable" performance levels for a safety system and provides requirements for evaluating the performance of a safety system. The ability to establish measurable performance levels allows to lower risk to an acceptable level [6].

Design of a SIS starts with Safety Life Cycle which covers all the SIS activities, right from initial conception to decommissioning, such as:

- Performing conceptual process design
- Performing Process Hazard Analysis & Risk Assessment
- Defining non-SIS protection layers
- Defining the need for an SIS
- Determining required Safety Integrity Level

ISA and IEC standards are based on the concept of safety life cycle, though there may be points where iterations are necessary.

Following are the some of design considerations, combination of which is used to meet the desired SIL of a SIS [7] .

A. Separation – Identical or Diverse

Separation between BPCS and SIS functions reduces the probability that both control and safety functions become unavailable at the same time, or that inadvertent changes affect the safety functions of the SIS. Therefore, it is generally necessary to provide separation between the BPCS and SIS functions.
Separation between the SIS and BPCS may be identical or diverse. Identical separation would mean using the same technology for both the BPCS and SIS whereas diverse separation would mean using different technologies for the same or different manufacturer.

Compared with identical separation, which helps against random failures, diverse separation offers the additional benefit of reducing the probability of systematic faults and of reducing common cause failures.

Identical separation between the SIS and BPCS may have some advantages in design and maintenance because it reduces the likelihood of maintenance errors. This is particularly the case if diverse components are to be selected, which have not been used before within the user's organization.

Following are the areas where separation between SIS and BPCS is needed to meet the safety functionality and safety integrity requirements:-

- Field sensors
- Final control elements
- Logic solver
- Wiring
- Communications between BPCS and SIS

Identical separation between SIS and BPCS is generally acceptable for SIL1 and SIL2 applications although the sources and effects of common cause failures should be considered and their likelihood reduced. For SIL3 safety instrumented functions, diverse separation is typically used to meet the required safety integrity.

On de-energize to trip systems, it is generally not necessary to separate the signals between the BPCS and SIS field instruments. This means the signals wires may be shared in a common multi-conductor cable and terminated in a common terminal box. Only for SIL1 application, use of single sensor/control valve is allowed, provided the safety integrity requirements are met.

There may be special case where it is not possible to provide separation between BPCS and SIS (e.g., a gas turbine control system includes both control and safety functions). Additional

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

considerations are required when combining control and safety functions in the same device. e.g.

- Evaluation of the failure of common components and software and their impact on SIS performance.
- Limiting access to the programming or configuration functions of the system.

B. Redundancy – Identical or Diverse

Redundancy can be applied to provide enhanced safety integrity or improved fault tolerance. The designer should determine the redundancy requirements that achieve the SIL and reliability requirements for all components of the SIS including sensors, logic solver and final control elements. It is applicable to both hardware and software. Diverse redundancy uses different technology, design, manufacture, software, firmware etc. to reduce the influence of common cause faults. Diverse technology should be used if it is required to meet the SIL. Diverse technology should not be used where its application can result in the use of lower reliability components that will not meet system reliability requirements. Some of the measures that can be used to achieve diverse redundancy are as follows:-

- The use of different measurement technologies of the same variable (e.g. displacer and differential pressure level transmitter)
- The use of different measurements (e.g. pressure and temperature) when there is a known relationship between them
- The use of geographic diversity (e.g. alternate routes for redundant communications media)
- The use of different types of PES for each channel of redundant architecture

C. Architecture

Selection of the SIS architecture is an activity performed during the conceptual design step of safety life cycle. The architecture has a major impact on the overall safety integrity and reliability of SIS. Some of the activities involved in determining the SIS architecture are as follows:-

- Selection of energize to trip or de-energize to trip design
- Selection of redundancy for power sources and SIS power supplies
- Selection of operator interface components (e.g. CRT, alarm annunciator, push-buttons) and their method of interconnection to the SIS
- Selection of data communication interface between SIS and other subsystems (e.g. BPCS) and their method of communication (e.g. read only or read/write)

Let us take an example. To meet the SIL3 requirements, SIS may include two separate and diverse 1oo1 (1 out of 1) arrangements, each with their own sensor, logic solver and final control element. The 1oo1 arrangements would be connected in a 1oo2 voting scheme. Diverse separation, redundancy and exhaustive diagnostic capabilities are considered significant aspects of a SIL3 systems.

D.SIS Management of Change (MOC)

The objective is to ensure that the MOC requirements are addressed in any changes made to an operating SIS. It requires a written procedure, which shall be in place to initiate, document, review, approve and implement any changes to an operating SIS. MOC procedure shall ensure that the following considerations are addressed prior to any change:-

- The technical basis and impact of proposed change on safety and health
- Authorization requirements for the proposed change
- Availability of memory space and effect on response time
- On-line versus off-line change
- Modification for operating procedures

Safety integrity level is also affected by the following parameters:-

- Device integrity (i.e. failure rate and failure mode)
- Functional testing interval (i.e. at a specific time interval, testing is performed to determine that the device can achieve the failsafe condition)
- Diagnostic coverage (i.e. automatic, on-line testing of various failure modes of a device)

III. ROLE OF QUANTITATIVE RELIABILITY ANALYSIS

Terms such as safety, reliability and availability are in a certain way connected with each other. In fact, various techniques that are applied in the field of reliability engineering are also applied for the determination of safety integrity levels. To prevent abnormal operating conditions from developing into an accident, high reliability of SIS is very important. Reliability and availability of SIS is linked to the estimation and evaluation of failure rates, failure modes and common cause failures of its components. Quantitative reliability analysis of safety instrumented systems represents a systematic tool for design optimization so as to strike a balance of safety, production, availability and cost. To perform the reliability calculations and to quantify the results, reliability data related to SIS subsystems is required. There are many sources of required reliability data e.g. end user (E&P companies) maintenance records, documented reliability studies, manufacturer data and public available data like OREDA (Offshore Reliability Database) or WOAD (Worldwide Offshore Accident Database) which are used for SIL determination and SIS design. Although generic data

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

represent the broad spectrum of failure modes/ failure rates across industry, yet its suitability and relevance for Indian offshore industry needs to be investigated. e.g.

Are shelf-state failure data from the vendors which is based on laboratory testing on predictive failures models include the impact of process environment?

Are failure data from valves used in North Sea representative for valves on Mumbai High offshore installation?

Are Indian offshore operation and maintenance practices which have direct impact on failure rates and failure modes are comparable with the operation and maintenance practices of Norway?

Several such issues associated with generic as well as vendor data, when used for safety instrumented systems for the Indian offshore oil & gas industry need to be answered by developing company specific failure data from all the offshore operating companies and integrating them in one common database [8].

A. Approach To Reliability Analysis

Some of the steps used to perform the reliability analysis of a typical Safety Instrumented System are as follows [9]:-

1) Development of methodology for performing Safety Integrity Level (SIL):

Within the regulatory framework of country and self defined acceptable risk criteria, companies use various methodologies to determine the target safety integrity level (SIL) of safety instrumented functions of safety instrumented system (SIS). Based on present regulatory requirements in India for offshore operations and resources committed by the company for the risk management, best suited methodology should be developed for SIL determination for target offshore installation of present study. Current standards, regulatory guidelines, design, operational & maintenance practices of safety instrumented systems (SIS) for production platforms operating in Indian offshore should be scrutinized to gain a clear understanding of current status. Previous SIL & reliability studies and safety audits carried out by the organizations should be reviewed and their findings should be critically analyzed. To record and measure the opinions of industry experts, questionnaires should be prepared along with interviews with corporate QHSE representatives, plant instrument engineers, design engineers and technical experts from suppliers of SIS components.

2) Development of methodology for collection and compilation of company specific failure frequency database:

Available failure frequency database like OREDA (Offshore Reliability Database) which are used presently for SIL determination and SIS design are generic in nature with almost negligible contribution from Indian Offshore Industry. Vendor supplied failure data is also uncertain as the failure modes/ failure rates of components of SIS are to large extent depend upon the company policies and actual process conditions [10]. A methodology should be developed for collection and compilation of company specific failure frequency database from offshore installations. To develop the company specific failure frequency database, a format should be designed to collect the data from offshore installation. Visit to offshore installations should be planned to collect archival records and history of operating safety instrumented systems. Format may have the provision to collect random failures, systematic failures, common cause failures, dangerous as well as safe failures and spurious trip failures. Vendor supplied failure data along with data related to diagnostic coverage and functional testing intervals should also be collected and compared with site specific data.

3) Calculation of reliability in terms of probability of failure on demand (PFD)

Reliability of various safety instrumented functions of safety instrumented system (SIS) is established in terms of average probability of failure of SIS on demand (PFDavg.). PFDavg. is calculated for each safety instrumented function of SIS using company specific failure data after applying suitable correction factors. Calculated values of reliability of safety instrumented functions should be used to verify the safety and reliability requirements of the offshore installation.

4) Study the factors affecting the result of reliability of target Safety Instrumented System (SIS)

Factors causing under-protection or over-protection of safety instrumented functions of target safety instrumented system (SIS) should be critically investigated after studying the existing design, implementation, operational and maintenance practices of target SIS. Based on the reliability evaluation of safety instrumented system and analysis of factors affecting it, specific recommendations should be brought forward to improve the reliability and overall performance of safety instrumented system of offshore oil & gas installation.

IV. CONCLUSION

It is currently observed that there is a growing need in offshore oil & gas industry to gain insight into the significant aspects and parameters of safety instrumented systems so as to manage the process in a more reliable and safer manner. Indian Exploration & Production (E&P) companies are currently struggling with uncertainty in reliability of safety instrumented systems due to a number of problems related to design, implementation, operation and maintenance of safety instrumented systems. A systematic quantitative reliability analysis can address, evaluate and resolve these concerning issues, which shall help the Indian E&P companies in more effective risk management of their offshore operations. This shall not only result in increased safety but also help the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

company to be more productive and effective in operational and maintenance practices, thus minimizing process downtime to the extent possible. The diversity of issues and the use of different subsystems demand a multidisciplinary team with expertise in process, instrumentation, control, safety, maintenance, reliability and management to develop the basis for the design, implementation, maintenance and finally the periodic quantitative reliability assessment of a SIS capable of achieving SIL requirements of high risk offshore oil and gas platforms.

REFERENCES

[1] ANSI/ISA-ISA 84.01-1996, ISA, Research Triangle Park, NC (1996): Application of Safety Instrumented Systems for the Process Industries.

[2] International Electro technical Commission (IEC), Ganeva (2003): IEC 61511: Functional Safety – safety instrumented systems for the process industry

[3] API (American Petroleum Institute) Recommended Practice (RP) 14C: Analysis, Design, Installation and Testing of Basic Surface Safety Systems on Offshore Production Platforms.

[4] API (American Petroleum Institute) Recommended Practice (RP) 14G: Recommended Practice for Fire Prevention and Control on Open Type Offshore Production Platforms.

[5] API (American Petroleum Institute) Recommended Practice (RP) 14J: Recommended Practice for Design and Hazard Analysis for Offshore Production Facilities.

[6] Wayne Ruschel, "The Future of Offshore Instrumented System," EDG Engineering, 2005 OREDA (1992).

[7] Rakesh Sethi, "Critical evaluation of selection criteria for safety instrumented system at offshore oil and gas offshore platforms," HSE Conference-2006, IPSHEM, 2006

[8] IEOT/RRE/2006-07(2006): HAZID/HAZOP studies in offshore/onshore construction.

[9] Rakesh Sethi, "Evaluation of reliability of safety instrumented system for risk management of offshore oil & gas production platforms in India." Punjabi University, Patiala, 2007

[10] Wang Y, West H.H, Mannan M.S. , "The impact of Data Uncertainty in determining Safety Integrity Level," Process Safety and Environmental Protection, 82 : 393-397 , 2004

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

PAPR REDUCTION OF OFDM SIGNAL USING ERROR CORRECTING CODE AND SELECTIVE MAPPING

Anshu
ECE Deptartment
Mahrishi Markendeshwar University
Mullana, India
guddal88@gmail.com

Er. Himanshu Sharma
Lecturer, ECE Department
Maharishi Markendeshwar University
Mullana, India
himanshu.zte@gmail.com

I. ABSTRACT

Orthogonal frequency division multiplexing (OFDM) technique is a promising technique to offer high data rate and reliable communications over fading channels. The main implementation disadvantage of OFDM is the possibility of high peak to average power ratio (PAPR). This paper presents a novel technique to reduce the PAPR using errorcorrecting coding and selective mapping (SLM).We show that the probability of the PAPR of OFDM signal with 100 subcarriers.

Keywords-OFDM,SLM,CCDF, PAPR, PAR.

II. INTRODUCTION

OFDM, orthogonal frequency division multiplexing, is a multicarrier communication technique, where a single data stream is transmitted over a number of lower rate subcarriers. OFDM has become tangible reality, it has been employed for wire-line communications and also has been employed in wireless local area network (WLAN) e.g. IEEE 802.11. Other applications of OFDM are digital audio broadcasting (DAB) and digital video broadcasting (DVB).

Unfortunately, OFDM has the drawback of a potentially high peak to average power ratio (PAPR). Since a multicarrier signal consists of a number of independent modulated subcarriers that can cause a large PAPR when the subcarriers are added up coherently.

To reduce the PAPR different techniques were proposed. These techniques can be categorized into the following, clipping and filtering [1], coding [2], phasing [3], scrambling [4], interleaving [5], and companding [6].

In this paper we propose and examine a technique for reducing the probability of a high PAPR, based on part on a method proposed in [1] and [8]. This technique is a variation of selective mapping (SLM) [1], in which a set of independent sequences are generated by some means from the original signal, and then the sequence with the lowest PAPR is transmitted. To generate these sequences we use code en encoder. Using error correcting coding will offer two advantages,significant PAPR reduction and astonishing bit error rate (BER) performance.

The rest of the paper is organized as follows: The problem of high PAPR of OFDM signal is briefly defined in section 2. Section 3 introduces the proposed technique. Some simulation results are shown in section 4. Finally, the conclusions are drawn in section 5.

III. PROBLEM DEFINITION

We suppose an OFDM transmission scheme, where a block of N complex symbols is first over-sampled using over sampling factor J and then transformed into time domain using the inverse fast Fourier transform (IFFT). This results in the following signal:

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k \, e^{\left(j2\pi \frac{kt}{\sqrt{N}}\right)} \qquad (1)$$

Where sk is the data to be transmitted and 1 < t < NJ.

The PAPR is defined as the ratio between the maximum powers occurring in OFDM symbol to the average power of the same OFDM symbol:

$$PAPR = \frac{max|x(t)|^2}{E[|x(t)^2|]} \qquad (2)$$

Where*E* [.] denotes expectation.
When the OFDM signal with high PAPR passes through a non-linear device, (power amplifier working in the saturation

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

region), the signal will suffer significant non-linear distortion [9]. This non-linear distortion will result in in-band distortion and out-of-band radiation. The in-band distortion causes system performance degradation and the out-of-band radiation causes adjacent channel interference (ACI) that affects systems working in the neighbour bands. To lessen the signal distortion, it requires a linear power amplifier with large dynamicrange. However, this linear power amplifier has poor efficiency and is so expensive.

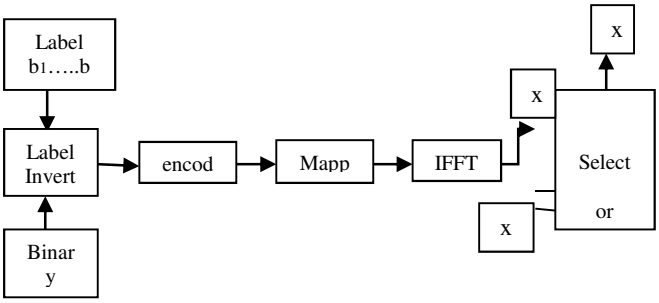


Figure: 1 System model

IV. SELECTIVE MAPPING USING ERROR CORRECTING CODING

Selected mapping (SLM) is a specific scheme for PAPRreduction that was introduced in [13].SLM takes advantage of the fact that the PAR of an OFDM signal is very sensitiveto phase shifts in the frequency-domain data. PAR reduction is achieved by multiplying independent phase sequences to the original data and determining the PAR of each phasesequence /data combination. The combination with the lowest PAR is transmitted. In other words, the data sequence X is element-wise phased by D N-length phase sequences,

$$\{\emptyset[k]^{(d)}\}_{k=o}^{N-1} = \emptyset^{(d)} \qquad (3)$$

Where d is an integer such that$d \in [0, D-1]$After phasing, the D possible frequency-domain OFDM symbols$X^{(d)} = X.e^{i\emptyset^{(d)}}$where • is the element-wise multiplication.

We assume that

$$\phi^{(0)} = 0 \qquad (4)$$

So that$X^{(0)} = X$ (5)

Define the D candidate time-domain OFDM symbols *x(d) = IFFT{X(d)}*. Note that all of the candidate symbols carry the same information. In order to achieve a PAPR reduction, the symbol with the lowest PAPR is transmitted. We define

$$\widetilde{d} = arg\ min\ PAPR\left(x^{(d)}\right) \qquad (6)$$
$$1 \le d \le D_{max}$$

In SLM, we use log2 D bits side information to indicate the phase weighting. As this side information is of highest importance to recover the data, it should be carefully protected by channel coding.

With$\tilde{d}$, the transmitted signal is$x^{(\tilde{d})}$. In the receiver, X can be recovered with

$$X\ FFT\{x^{(\widetilde{(d)})}\} \cdot e^{-j\emptyset(\tilde{d})} \qquad (7)$$
$$X \cdot e^{j\emptyset(\widetilde{d})} \cdot\ e^{-j\emptyset(\widetilde{d})}\ X$$

To recover X it is necessary for the receiver to have a table of all$\phi(d)$.

The phase sequence combination that results in the lowest PAR of the time-domain signal is used for transmission. Here we encode the information fist with forward error correcting code and then do SLM. The technology combines SLM, which aims at PAPR reduction, and coding(basically we use hamming code, rsc code, and convolution code), which are excellent in error control and are play excellent role in further reduction in PAPR. In this paper we show comparison of all the three coding with SLM.

After coding the process is done just like what is applied on the carriers in the usual SLM algorithm. The process is given in figure1. Finally all of the different sequences, after serial to parallel conversion, pass through the IFFT block to produce D block of time domain signal. The block with the lowest PAPR is to be sent to the receiver through the channel.

V. SIMULATION RESULTS

In this section the results obtain through the simulations using MATLAB are examined. Quarter phase shift keying (QPSK) is used. The results are given here in terms of PAPR-CCDF. First we show comparison of PAPR with SLM and the original OFDM signal. Secondly the comparison with SLM and coded SLM.

In the following the performance of the system for both conventional SLM and coded are examined.

A. CCDF-PAPR

Figure 1 shows the difference between the results of SLM and Original OFDM signal in terms of CCDF-PAPR. In this we consider the OFDM signal with number of subcarriers N=64. The hamming encoder has constraint length K=16 is used.

B. CCDF-PAPR

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 2 in this section, the performance of the system for each coded SLM i.e. SLM using hamming code,SLM using rsc code and SLM using convolution code. In this the OFDM signal with number of subcarriers N=100 and hamming encoder has constraint lengthK=16 is used.

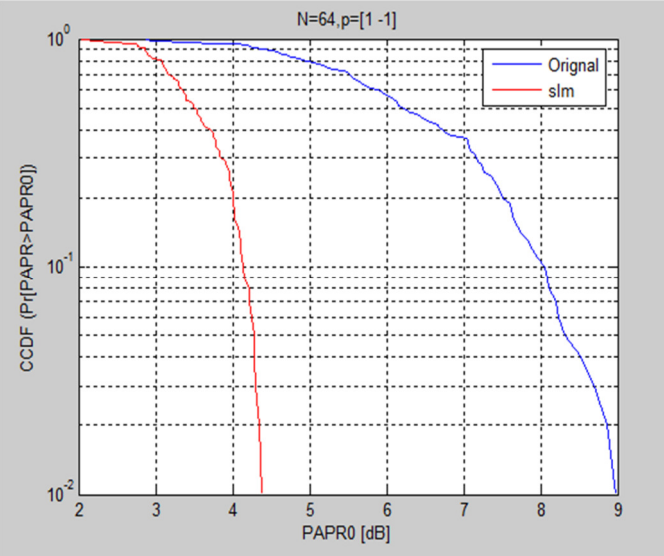


Figure2: Comparison between the original OFDM &SLM

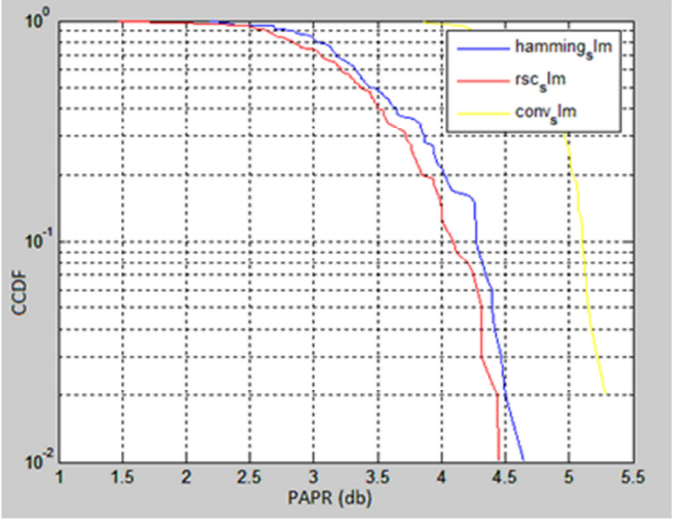


Figure3: Comparison among the various coding with SLM

VI. CONCLUSIONS

We have shown that, coding and SLM can be combined to reduce the PAPR of OFDM signal with quite moderate additional complexity. The advantage of the proposed scheme is that, the coding is used for two purposes, error correction and PAPR reduction. Here we see that combination of RS code and SLM gives the better results instead of hamming with SLM and convolution with SLM as shown in fig 2.This reduces the hardware complexity of the system. An excellent style manual for science writers is [7].

VII. ACKNOWLEDGEMENT

I would like to express my gratitude toward those who have supported me through the work. First, I would like give my sincerest thanks to my guide, Er. Himanshu Sharma, for his unreserved guidance, support and encouragement. He has looked out for me at every time and has given me every opportunity to succeed; for that, I am truly grateful. I consider myself lucky to be one of her students. I want to thank my thesis committee, Dr.AmitGarg and Dr. H.P Sinha, for taking the time to review and critique my work. Additionally, I would like to thank my fellow research group members for their friendship and insightful comments. Finally, I want to thank my family, who have supported me through every step of my life.

VIII. REFERENCES

[1] X. Li and L. J. Cimini Jr., ”Effects of clipping and filtering on the performance of OFDM,” IEEE Communication Letters, pp. 131-133, May 1998.

[2] A. E. Jones, T. A. Wilkinson and S. K. Barton, ”Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes,” Electronic Letters, pp. 2098-2099, Dec. 1994.

[3] V. Tarokh and H. Jafakhani, ”On the computation and reduction of peak to average power ratio in multicarrier communications,” Proc. IEEE 53rd Vehicular Technology Conference, vol.2, pp. 37-44, Jan 2000.

[4] P. Van Eetvelt, G. Wade and M. Tomlinson, ”Peak to average power reduction for OFDM schemes by selective scrambling,” Electronic letters, vol. 32, pp. 1963-1994, Jan 2000.

[5] A. D. S. Jayalath and C. Tellambura ” The use of interleaving to reduce the peak to average power ratio of an OFDM signal,” Global Telecommunication

Conference, 2000, GLOBECOM ’00. IEEE, vol. 1, pp. 82-86, Nov.-1 Dec. 2000.

[6] Xiao Huang, Jianhua Lu, Justin Chuang, and JunliZheng, ”Companding transform for the reduction of peak to average power ratio of OFDM signal,” IEEE Trans. On Common. vol. 48, pp. 835-839, May 2001.

[7] R. B¨auml, R. Fischer and J. Huber, ”Reducing the peak to average power ratio of multicarrier modulation by selected mapping,” Elect. Letts., vol. 32, pp. 2056-2057, Oct. 1996.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[8] N. Carson, T. A. Gulliver, ”PAPR reduction of OFDM using selected mapping, modified RA codes and clipping”, in Proc. IEEE VTC, vol.2, pp. 1070-1073, Sep. 2002.

[9] S. Merchan, A. Garcia Armada and J. L. Garcia,” OFDM performance in ampli- er nonlinearity,” IEEE Trans. On Broadcasting, vol. 44 Issue:1, pp.106-114, Mar. 1998.

[10] M. Breiling, S. M¨uller-Weinfurtner and J. Huber, ”Peak-power reduction in OFDM without explicit side information,” Proc. 5th Int. OFDM Workshop, Germany, pp- 28.1-28.4,2000.

[11] C. Berrou, A. Glavieux and P. Thitimajshima, ”Near Shannon limit error correcting coding and decoding: Turbo Codes.” in Proc. of ICC’93, (Geneva, Switzerland), pp. 1064-1070, May 1993.

[12] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, ”A soft-input soft-output maximum a posterior (MAP) module to decode parallel and serial concatenated codes”, TDA progress report 42-127, Jet Propulsion Lab., Pasadena, CA, Nov. 15 1996.

[13] K.Kasiri&M.j.Dehghani, “A Blind SLM Scheme for Reduction of PAPR in OFDM System.”

Author Profile:

Anshu
M.Sc., Kurukshetra University
Pursuing M.Tech at Mahrishi Markendeshwar University, MULLANA

Er. Himanshu Sharma
M.E.(ECE)from THAPER UNIVERSITY, PATIALA
Currently he is a Lecturer, ECE Department
Mahrishi Markendeshwar University, MULLANA
Experience in Wireless Communication.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Lossless Image Compression For Transmitting Over Low Bandwidth Line

G. Murugan, Research Scholar , Singhania University & Sri Venkateswara College of Engg , Thiruvallur

Dr. E. Kannan, Supervisor ,Singhania University and Dean Academic Veltech University

S. Arun , ECE Dept. Asst.Professor Veltech High Engg college,Chennai email-yesarun001@yahoo.com

Abstract

The aim of this paper is to develop an effective loss less algorithm technique to convert original image into a compressed one. Here we are using a lossless algorithm technique in order to convert original image into compressed one. Without changing the clarity of the original image. Lossless image compression is a class of image compression algorithms that allows the exact original image to be reconstructed from the compressed data.

We present a compression technique that provides progressive transmission as well as lossless and near-lossless compression in a single framework. The proposed technique produces a bit stream that results in a progressive and ultimately lossless reconstruction of an image similar to what one can obtain with a reversible wavelet codec. In addition, the proposed scheme provides near-lossless reconstruction with respect to a given bound after decoding of each layer of the successively refineable bit stream. We formulate the image data compression problem as one of successively refining the probability density function (pdf) estimate of each pixel. Experimental results for both lossless and near-lossless cases indicate that the proposed compression scheme, that innovatively combines lossless, near-lossless and progressive coding attributes, gives competitive performance in comparison to state-of-the-art compression schemes.

1.INTRODUCTION

Lossless or reversible compression refers to compression techniques in which the reconstructed data exactly matches the original. Near-lossless compression denotes compression methods, which give quantitative bounds on the nature of the loss that is introduced. Such compression techniques provide the guarantee that no pixel difference between the original and the compressed image is above a given value [1]. Both lossless and near-lossless compression find potential applications in remote sensing, medical and space imaging, and multispectral image archiving. In these applications the volume of the data would call for lossy compression for practical storage or transmission. However, the necessity to preserve the validity and precision of data for subsequent reconnaissance diagnosis operations, forensic analysis, as well as scientific or clinical measurements, often imposes strict constraints on the reconstruction error. In such situations near-lossless compression becomes a viable solution, as, on the one hand, it provides significantly higher compression gains vis-à-vis lossless algorithms, and on the other hand it provides guaranteed bounds on the nature of loss introduced by compression.

Another way to deal with the lossy-lossless dilemma faced in applications such as medical imaging and remote sensing is to use a successively refindable compression technique that provides a bit stream that leads to a progressive reconstruction of the image. Using wavelets, for example, one can obtain an embedded bit stream from which various levels of rate and distortion can be obtained. In fact with reversible integer wavelets, one gets a progressive reconstruction capability all the way to lossless recovery of the original. Such techniques have been explored for potential use in tele-radiology where a physician typically requests portions of an image at increased quality (including lossless reconstruction) while accepting initial renderings and unimportant portions at lower quality, and thus reducing the overall bandwidth requirements. In fact, the new still image compression standard, JPEG 2000, provides such features in its extended form [2].

In this paper, we present a compression technique that incorporates the above two desirable characteristics, namely, near-lossless compression and progressive refinement from lossy to lossless reconstruction. In other words, the proposed technique produces a bit stream that results in a progressive reconstruction of the image similar to what one can obtain with a reversible wavelet codec. In addition, our scheme provides near-lossless (and lossless) reconstruction with respect to a given bound after each layer of the successively refinable bit stream is decoded. Note, however that these bounds need to be set at compression time and cannot be changed during decompression. The compression performance provided by the proposed technique is comparable to the best-known lossless and near-lossless techniques proposed in the literature. It should be noted that to the best knowledge of the authors, this is the first technique reported in the literature that provides lossless and near-lossless compression as well as progressive reconstruction all in a single framework.

2. METHODOLOGY

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2.1COMPRESSION TECHNIQUES

- **LOSSLESS COMPRESSION**

Where data is compressed and can be reconstituted (uncompressed) without loss of detail or information. These are referred to as bit-preserving or reversible compression systems also [11].

- **LOSSY COMPRESSION**

Where the aim is to obtain the best possible fidelity for a given bit-rate or minimizing the bit-rate to achieve a given fidelity measure. Video and audio compression techniques are most suited to this form of compression [12].

- If an image is compressed it clearly needs to be uncompressed (decoded) before it can viewed/listened to. Some processing of data may be possible in encoded form however.
- Lossless compression frequently involves some form of ***entropy encoding*** and are based in information theoretic techniques
- Lossy compression use source encoding techniques that may involve transform encoding, differential encoding or vector quantisation

Image compression may be lossy or lossless. Lossless compression is preferred for archival purposes and often for medical imaging, technical drawings, clip art, or comics. This is because lossy compression methods, especially when used at low bit rates, introduce compression artifacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate. The lossy compression that produces imperceptible differences may be called visually lossless.

2.2METHODS FOR LOSSLESS IMAGE COMPRESSION ARE

- Run-length encoding – used as default method in PCX and as one of possible in BMP, TGA, TIFF
- DPCM and Predictive Coding
- Entropy encoding
- Adaptive dictionary algorithms such as LZW – used in GIF and TIFF
- Deflation – used in PNG, MNG, and TIFF
- Chain codes

2.3METHODS FOR LOSSY COMPRESSION

- Reducing the color space to the most common colors in the image. The selected colors are specified in the color palette in the header of the compressed image. Each pixel just references the index of a color in the color palette. This method can be combined with dithering to avoid posterization.
- Chroma sub sampling. This takes advantage of the fact that the human eye perceives spatial changes of brightness more sharply than those of color, by averaging or dropping some of the chrominance information in the image.
- Transform coding. This is the most commonly used method. A Fourier-related transform such as DCT or the wavelet transform are applied, followed by quantization and entropy coding.
- Fractal compression.

2.3COMPRESSION

The process of coding that will effectively reduce the total number of bits needed to represent certain information.

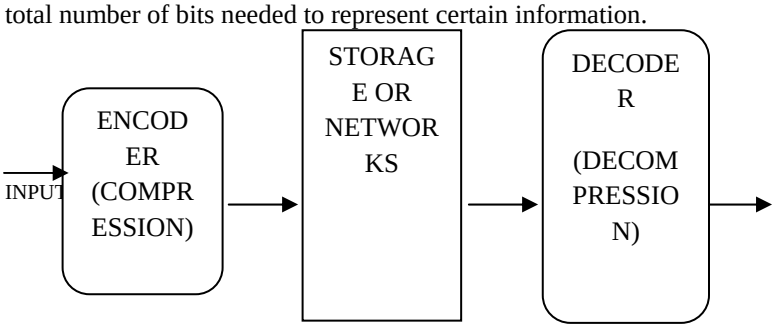


Fig.1. a general data compression scheme

Fig.2 lossy image compressionresult result

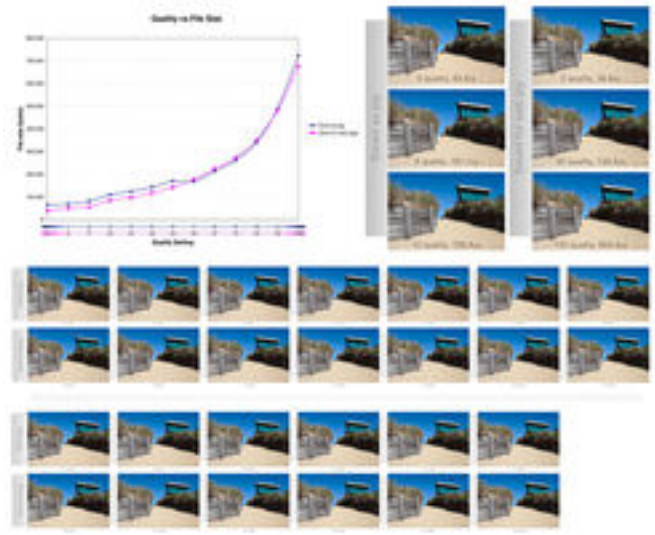
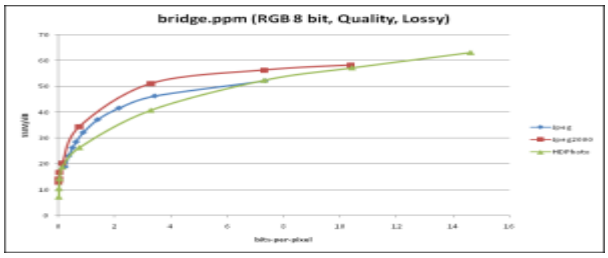


Fig. 3 lossless image comparison ratio

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



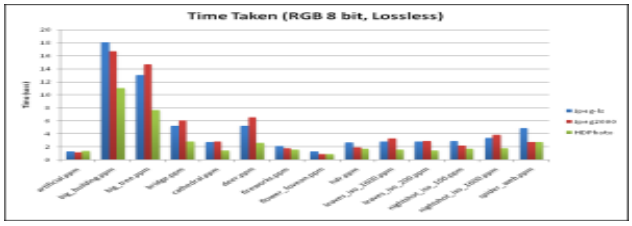


Fig.4lossy and lossless comparison ratio

3.HUFFMAN CODING

Huffman coding is based on the frequency of occurrence of a data item (pixel in images). The principle is to use a lower number of bits to encode the data that occurs more frequently. Codes are stored in a Code Book which may be constructed for each image or a set of images. In all cases the code book plus encoded data must be transmitted to enable decoding.

The Huffman algorithm is now briefly summarised:

- A bottom-up approach
- 1. Initialization: Put all nodes in an OPEN list, keep it sorted at all times (e.g., ABCDE).
- 2. Repeat until the OPEN list has only one node left:
- (a) From OPEN pick two nodes having the lowest frequencies/probabilities, create a parent node of them.
- (b) Assign the sum of the children's frequencies/ probabilities to the parent node and insert it into OPEN.
- (c) Assign code 0, 1 to the two branches of the tree, and delete the children from OPEN.

The following points are worth noting about the above algorithm:

Decoding for the above two algorithms is trivial as long as the coding table (the statistics) is sent before the data. (There is a bit overhead for sending this, negligible if the data file is big.)

Unique Prefix Property

No code is a prefix to any other code (all symbols are at the leaf nodes) great for decoder, unambiguous. If prior statistics are available and accurate, then Huffman coding is very good.

3.1HUFFMAN CODING OF IMAGES

In order to encode images:

- Divide image up into 8x8 blocks
- Each block is a symbol to be coded
- Compute Huffman codes for set of block
- Encode blocks accordingly

3.2HUFFMAN CODING ALGORITHM

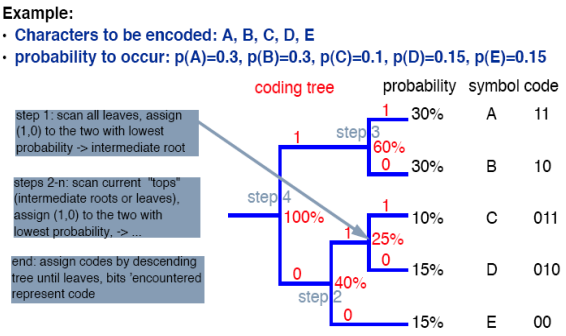


Table and example of application to data stream

symbol	code
A	11
B	10
C	011
D	010
E	00

B A C D A B E B A E

10 11 011 010 11 10 00 10 11 00

No Huffman code is the prefix of any other Huffman codes so decoding is unambiguous

- The Huffman coding technique is optimal (but we must know the probabilities of each symbol for this to be true)
- Symbols that occur more frequently have shorter Huffman codes

4.LEMPEL-ZIV-WELCH (LZW) ALGORITHM

THE LZW COMPRESSION ALGORITHM CAN SUMMARISED AS FOLLOWS

```
w = NIL;
while ( read a character k )
    {
            if wk exists in the dictionary
                    w = wk;
            else
                    add wk to the dictionary;
                    output the code for w;
                    w = k;
    }
```

THE LZW DECOMPRESSION ALGORITHM IS AS FOLLOWS

```
read a character k;
    output k;
    w = k;
    while ( read a character k )
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
/* k could be a character or a code. */
        {
                entry = dictionary entry for k;
                output entry;
                add w + entry[0] to dictionary;
                w = entry;

        }
```

4.2ENTROPY ENCODING

- Huffman maps fixed length symbols to variable length codes. Optimal only when symbol probabilities are powers of 2.
- Arithmetic maps entire message to real number range based on statistics. Theoretically optimal for long messages, but optimality depends on data model. Also can be CPU/memory intensive.
- Lempel-Ziv-Welch is a dictionary-based compression method. It maps a variable number of symbols to a fixed length code.
- Adaptive algorithms do not need a priori estimation of probabilities, they are more useful in real applications.

4.2.1LOSSLESS JPEG

- JPEG offers both lossy (common) and lossless (uncommon) modes.
- Lossless mode is much different than lossy (and also gives much worse results)
- Added to JPEG standard for completeness
- Lossless JPEG employs a predictive method combined with entropy coding.
- The prediction for the value of a pixel (greyscale or color component) is based on the value of up to three neighboring pixels
- One of 7 predictors is used (choose the one which gives the best result for this pixel).

PREDICTOR	PREDICTION
P1	A
P2	B
P3	C
P4	A+B-C
P5	A+(B-C)/2
P6	B+(A-C)/2
P7	(A+B)/2

Table lossless jpeg

- Now code the pixel as the pair (predictor-used, difference from predicted method)
- Code this pair using a lossless method such as Huffman coding
  - The difference is usually small so entropy coding gives good results
  - Can only use a limited number of methods on the edges of the image

5.LOSSY AND LOSSLESS ALGORITHMS

TREC includes both lossy and lossless compression algorithms. The lossless algorithm is used to compress data for the Windows desktop which needs to be reproduced exactly as it's decompressed. The lossy algorithm is used to compress 3D image and texture data when some loss of detail is tolerable.

Let me just explain the point about the Windows desktop since it's perhaps not obvious why I even mentioned it. A Talisman video card in a PC is not only going to be producing 3D scenes but also the usual desktop for a Windows platform. Since there is no frame buffer, the entire desktop needs to be treated as a sprite which in effect forms a background scene on which 3D windows might be superimposed. Obviously we want to use as little memory as possible to store the Windows desktop image so it makes sense to try to compress it, but it's also vital that we don't distort any of the pixel data since it is possible that an application might want to read back a pixel it just wrote to the display via GDI. So some form of lossless algorithm is vital when compressing the desktop image.

5.1LOSSLESS COMPRESSION

Let's take a look at how the lossless compression algorithm works first as it the simpler of the two. Figure 4.1 shows a block diagram of the compression process.

Fig. 4.1 the lossless compression process

The RGB data is first converted to a form of YUV. Using a YUV color space instead of RGB provides for better compression. The actual YUV data is peculiar to the TREC algorithm and is derived as follows:

$$Y = G$$

$$U = R - G$$

$$V = B - G$$

The conversion step from RGB to YUV is optional. Following YUV conversion is a prediction step which takes advantage of the fact that an image such as a typical Windows desktop has a lot of vertical and horizontal lines as well as large areas of solid color. Prediction is applied to each of the R, G, B and alpha values separately. For a given pixel p(x, y) it's predicted value d(x, y) is given by

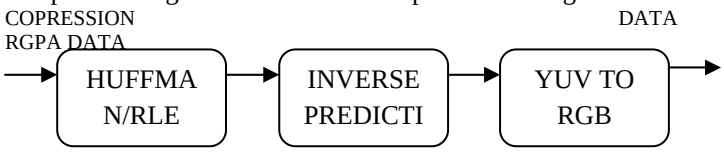
http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$d(0, 0) = p(0, 0)$$
$$d(0, y) = p(0, y) - p(0, y-1) \quad \text{for } y > 0$$
$$d(x, y) = p(x, y) - p(x-1, y) \quad \text{for } x > 0$$

The output values from the predictor are fed into a Huffman/RLE encoder which uses a set of fixed code tables. The encoding algorithm is the same as that used in JPEG for encoding the AC coefficients. (See ISO International Standard 10918, " Digital Compression and Coding of Continuous-Tone Still Images".) The Huffman/RLE encode outputs a series of variable-length code words. These code words describe the length from 0 to 15 of a run of zeroes before the next coefficient and the number of additional bits required to specify the sign and mantissa of the next non-zero coefficient. The sign and mantissa of the non-zero coefficient then follow the code word.

5.2LOSSLESS DECOMPRESSION

Decompressing an image produced by the lossless compression algorithm follows the steps shown in figure 4.2



5.2.1the lossless decompression process

The encoded data is first decoded using a Huffman decoder using fixed code tables. The data from the Huffman decoder is then passed through the inverse of the prediction filter used in compression. For predicted pixel d(x, y) the output pixel values p(x, y) are given by:

$$p(0, 0) = d(0, 0), p(0, y) = d(0, y-1) + d(0, y) \quad \text{for } y > 0$$
$$p(x, y) = d(x-1, y) + d(x, y) \quad \text{for } x > 0$$

The final step is to convert the YUV-like data back to RGB using: R = Y + U, G = Y,B = Y + V

5.3LOSSY COMPRESSION

The lossy compression algorithm is perhaps more interesting since it achieves much higher degrees of compression that the lossless algorithm and is used more extensively in compressing the 3D images we are interested in. Figure 3 shows the compression steps.

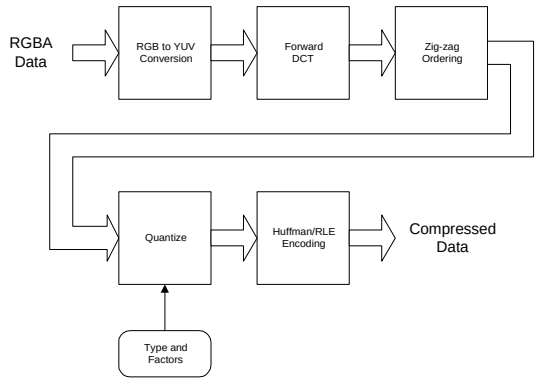


Fig. 4.3 the lossy compression process

The first step is to convert the RGB data to a form of YUV called YOrtho using the following:

$$Y = (4R + 4G + 4B) / 3 - 512$$
$$U = R - G$$
$$V = (4B - 2R - 2G) / 3$$

Note that the alpha value is not altered by this step. The next step is to apply a two-dimensional Discrete Cosine Transform (DCT) to each color and alpha component. This produces a two-dimensional array of coefficients for a frequency domain representation of each color and alpha component. The next step is to rearrange the order of the coefficients so that low DCT frequencies tend to occur at low positions in a linear array. This tends to place zero coefficients in the upper end of the array and has the effect of simplifying the following quantization step and improving compression through the Huffman stage. The quantization step reduces the number of possible DCT coefficient values by doing an integer divide. Higher frequencies are divided by higher factors because the eye is less sensitive to quantization noise in the higher frequencies. The quantization factor can vary from 2 to 4096. Using a factor of 4096 produces zeros for all input values. Each color and alpha plane has its own quantization factor. Reducing the detail in the frequency domain by quantization leads to better compression and the expense of lost detail in the image. The quantized data is then Huffman encoded using the same process as was described for lossless compression.

5.4LOSSY DECOMPRESSION

The decompression process for images compressed using the TREC lossy compression algorithm is shown in figure

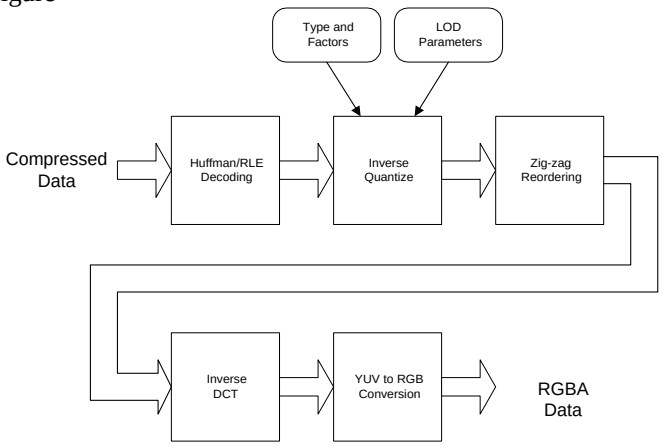


Fig. 4.4 the lossy decompression process

The decompression process is essentially the reverse of that used for compression except for the inverse quantization stage. At this point a level of detail (LOD) parameter can be used to determine how much detail is required in the output image. Applying a LOD filter during decompression is useful when reducing the size of an image. The LOD filter removes the higher frequency DCT coefficients which helps avoid aliasing in the output image

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

when simple pixel sampling is being used to access the source pixels.

Note that the level of detail filtering is not a part of the TREC specification and not all TREC decompressions will implement it.

6.EXPERIMENTAL RESULTS

We present experimental results based on the steps

Step1. Lossless Compression

Step2. Lossless Decompression

Step3. Lossless image Compression using Huffman coding

Step4. Lossless image Decompression using Huffman coding

Step5. Lossless image Compression for transmitting Low Bandwidth Line

7.CONCLUSIONS

This work has shown that the compression of image can be improved by considering spectral and temporal correlations as well as spatial redundancy. The efficiency of temporal prediction was found to be highly dependent on individual image sequences. Given the results from earlier work that found temporal prediction to be more useful for image, we can conclude that the relatively poor performance of temporal prediction, for some sequences, is due to spectral prediction being more efficient than temporal. Another Conclusions and Future Work finding from this work is that the extra compression available from image can be achieved without necessitating a large increase in decoder complexity. Indeed the presented scheme has a decoder that is less complex than many lossless image compression decoders, due mainly to the use of forward rather than backward adaptation.

Although this study considered a relatively large set of test image sequences compared to other such studies, more test sequences are needed to determine the extent of sequences for which temporal prediction is more efficient than spectral prediction..

8.REFERENCES

[1] N. Memon and K. Sayood. Lossless image compression: A comparative study. Proc. SPIE Still-Image Compression, 2418:8–20, March 1995.

[2] N. Memon and K. Sayood. Lossless compression of rgb color images. Optical Engineering, 34(6):1711–1717, June 1995.

[3] S. Assche, W. Philips, and I. Lemahieu. Lossless compression of pre-press images using a novel color decorrelation technique. Proc. SPIE Very High Resolution and QualityImaging III, 3308:85–92, January 1998.

[4] N. Memon, X. Wu, V. Sippy, and G. Miller. An interband coding extension of the new lossless jpeg standard. Proc. SPIE Visual Communications and Image Processing, 3024:47–58, January 1997.

[5] N. Memon and K. Sayood. Lossless compression of video sequences. IEEE Trans. on. Communications, 44(10):1340–1345, October 1996.

[6] S. Martucci. Reversible compression of hdtv images using median adaptive prediction and arithmetic coding. Proc. IEEE International Symposium on Circuits and Systems, pages 1310–1313, 1990.

[7] M. Weinberger, G. Seroussi, and G. Sapiro. LOCO-I: A low complexity, context-based, lossless image compression algorithm. Proc. Data Compression Conference, pages 140–149, March 1996.

[8] ITU-T Rec. T.84 - Information Technology. Digital compression and coding of continuous-tone still images: Extensions, July 1996.

[9] M. Weinberger, J. Rissanen, and R. Arps. Applications of universal context modeling to lossless compression of gray-scale images. IEEE Trans. on Image Processing, 5(4):575–586, April 1996.

[10] G. Wallace. The jpeg still picture compression standard. Comms. of the ACM, 34(4):30– 44, April 1991.

[11] ISO/IEC 14495-1, ITU Recommendation T.87, “Information technology - Lossless and near-lossless compression of continuous-tone still images,” 1999.

[12] M. J. Weinberger, G. Seroussi, and G. Sapiro, “LOCO-I: A low complexity lossless image compression algorithm.” ISO/IEC JTC1/SC29/WG1 document N203, July 1995.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2011
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500